GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:52:54 ; Search time 55 Seconds
(without alignments)
128.431 Million cell updates/sec

Title: US-10-019-482-1
Perfect score: 105
Sequence: 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	96.2	25	4	AAB66787	Aab66787 Amphipath
2	65	61.9	104	7	ADE10685	Ade10685 Structura
3	65	61.9	104	7	ADE10635	Ade10635 Structura
4	62.5	59.5	428	6	ABU27824	Abu27824 Protein e
5	62	59.0	104	7	ADE10683	Ade10683 Structura
6	62	59.0	104	7	ADE10682	Ade10682 Structura
7	62	59.0	104	7	ADE10633	Ade10633 Structura
8	62	59.0	104	7	ADE10632	Ade10632 Structura
9	61.5	58.6	104	7	ADE10684	Ade10684 Structura
10	61.5	58.6	104	7	ADE10634	Ade10634 Structura
11	61	58.1	59	7	ADE10698	Ade10698 Structura
12	61	58.1	59	7	ADE10648	Ade10648 Structura
13	61	58.1	67	7	ADE10697	Ade10697 Structura
14	61	58.1	67	7	ADE10647	Ade10647 Structura
15	61	58.1	75	7	ADE10696	Ade10696 Structura
16	61	58.1	75	7	ADE10646	Ade10646 Structura
17	61	58.1	83	7	ADE10695	Ade10695 Structura
18	61	58.1	83	7	ADE10645	Ade10645 Structura
19	61	58.1	88	7	ADE10642	Ade10642 Structura
20	61	58.1	88	7	ADE10692	Ade10692 Structura
21	61	58.1	91	7	ADE10694	Ade10694 Structura
22	61	58.1	91	7	ADE10644	Ade10644 Structura
23	61	58.1	104	7	ADE10690	Ade10690 Structura
24	61	58.1	104	7	ADE10640	Ade10640 Structura
25	61	58.1	105	7	ADE10636	Ade10636 Structura
26	61	58.1	105	7	ADE10686	Ade10686 Structura
27	61	58.1	106	7	ADE10639	Ade10639 Structura
28	61	58.1	106	7	ADE10688	Ade10688 Structura
29	61	58.1	106	7	ADE10687	Ade10687 Structura
30	61	58.1	106	7	ADE10689	Ade10689 Structura
31	61	58.1	106	7	ADE10638	Ade10638 Structura
32	61	58.1	106	7	ADE10637	Ade10637 Structura
33	61	58.1	110	7	ADE10641	Ade10641 Structura
34	61	58.1	111	7	ADE10691	Ade10691 Structura
35	61	58.1	623	6	ABJ25843	Abj25843 Aspergill
36	61	58.1	700	6	ABJ26443	Abj26443 Aspergill
37	60	57.1	33	2	AAR90181	Aar90181 Polycatio
38	60	57.1	33	2	AAW06688	Aaw06688 Protamine
39	59.5	56.7	56	3	AAY82573	Aay82573 Copolymer
40	59	56.2	421	6	ABU28559	Abu28559 Protein e
41	58	55.2	78	7	ADE10666	Ade10666 Structura
42	58	55.2	78	7	ADE10616	Ade10616 Structura
43	58	55.2	104	7	ADE10681	Ade10681 Structura
44	58	55.2	104	7	ADE10680	Ade10680 Structura
45	58	55.2	104	7	ADE10631	Ade10631 Structura

ALIGNMENTS

RESULT 1
AAB66787
ID AAB66787 standard; peptide; 25 AA.
XX
AC AAB66787;
XX
DT 11-APR-2001 (first entry)
XX
DE Amphipathic peptide conjugate.
XX
KW Amphipathic; lipid bilayer; detergent.
XX
OS Synthetic.
XX
PN WO200102425-A2.
XX
PD 11-JAN-2001.
XX
PF 29-JUN-2000; 2000WO-CA000773.
XX
PR 29-JUN-1999; 99US-0140988P.
XX
PA (UYHE-) UNIV HEALTH NETWORK.
XX
PI Prive G;
XX
DR WPI; 2001-138120/14.
XX
PT New amphiphatic peptide conjugate having detergent properties, and
PT hydrophobic and hydrophilic phase, useful e.g. for stabilizing and
PT crystallizing proteins and membrane proteins, as cytolytic agents,
PT surfactants or emulsifiers.
XX
PS Claim 1; Page 22; 29pp; English.
XX
CC The present invention relates to an amphipathic peptide conjugate having
CC detergent properties and a hydrophobic and hydrophilic face. The
CC amphipathic peptide conjugate may be used for the stabilization and
CC crystallization of proteins and membrane proteins,for modifying the
CC properties of lipid bilayer membranes, as cytolytic agents, as molecules
CC that can facilitate the transport of polar molecules across biological
CC membranes, and as emulsifiers and surfactants
XX
SQ Sequence 25 AA;

Query Match 96.2%; Score 101; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              |||||||||||||||||||||||||
Db          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
```

RESULT 2
ADE10685
ID ADE10685 standard; protein; 104 AA.
XX
AC ADE10685;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library scaffold protein seqid 92.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response;
KW scaffold protein.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Disclosure; SEQ ID NO 92; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 104 AA;

```
  Query Match             61.9%;  Score 65;  DB 7;  Length 104;
  Best Local Similarity   68.0%;  Pred. No. 0.12;
  Matches   17;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| ||||| |||   |
Db         10 AAAAAAEAAAKAAAEAAAKAAAEAA 34
```

RESULT 3
ADE10635
ID ADE10635 standard; protein; 104 AA.
XX
AC ADE10635;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library related protein seqid 42.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Example 6; SEQ ID NO 42; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;

CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a protein associated with
CC fused nucleic acid and random peptide libraries of the invention.
XX
SQ Sequence 104 AA;

Query Match 61.9%; Score 65; DB 7; Length 104;
Best Local Similarity 68.0%; Pred. No. 0.12;
Matches 17; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
Db 10 AAAAAAEAAAKAAAEAAAKAAAEAA 34

RESULT 4
ABU27824
ID ABU27824 standard; protein; 428 AA.
XX
AC ABU27824;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #13351.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Enterobacter cloacae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA31694.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 55748; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 428 AA;

Query Match 59.5%; Score 62.5; DB 6; Length 428;
Best Local Similarity 64.3%; Pred. No. 1.3;
Matches 18; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

Qy 1 AXAEAAEKAA-----KYAAEAAEKAAKA 23
Db 210 AEAEAAKKAAQEAEKKAAAEAAKKAAAA 237

RESULT 5
ADE10683
ID ADE10683 standard; protein; 104 AA.
XX
AC ADE10683;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library scaffold protein seqid 90.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response;
KW scaffold protein.

XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Disclosure; SEQ ID NO 90; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 104 AA;

Query Match 59.0%; Score 62; DB 7; Length 104;
Best Local Similarity 72.0%; Pred. No. 0.31;
Matches 18; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

```
Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || |||  |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31
```

RESULT 6
ADE10682
ID ADE10682 standard; protein; 104 AA.
XX
AC ADE10682;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library scaffold protein seqid 89.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response;
KW scaffold protein.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Disclosure; SEQ ID NO 89; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug

CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 104 AA;

Query Match 59.0%; Score 62; DB 7; Length 104;
Best Local Similarity 72.0%; Pred. No. 0.31;
Matches 18; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
| |||| ||| || ||| |||| |
Db 9 AAAEAAAKAA--AAAAAEAAAKAAA 31

RESULT 7
ADE10633
ID ADE10633 standard; protein; 104 AA.
XX
AC ADE10633;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library related protein seqid 40.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Example 6; SEQ ID NO 40; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a protein associated with
CC fused nucleic acid and random peptide libraries of the invention.
XX
SQ Sequence 104 AA;

Query Match 59.0%; Score 62; DB 7; Length 104;
Best Local Similarity 72.0%; Pred. No. 0.31;
Matches 18; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
| |||| ||| || ||| |||| |
Db 9 AAAEAAAKAA--AAAAAEAAAKAAA 31

RESULT 8
ADE10632
ID ADE10632 standard; protein; 104 AA.
XX
AC ADE10632;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library related protein seqid 39.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX

DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Example 6; SEQ ID NO 39; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a protein associated with
CC fused nucleic acid and random peptide libraries of the invention.
XX
SQ Sequence 104 AA;

Query Match 59.0%; Score 62; DB 7; Length 104;
Best Local Similarity 72.0%; Pred. No. 0.31;
Matches 18; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
| |||| ||| || ||| |||| |
Db 9 AAAEAAAKAA--AAAAAEAAAKAAA 31

RESULT 9
ADE10684
ID ADE10684 standard; protein; 104 AA.
XX
AC ADE10684;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library scaffold protein seqid 91.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response;
KW scaffold protein.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Disclosure; SEQ ID NO 91; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 104 AA;

Query Match 58.6%; Score 61.5; DB 7; Length 104;
Best Local Similarity 72.0%; Pred. No. 0.37;
Matches 18; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

```
                  | | ||| ||| || ||| |||| |
Db          6 AAAAAAEAAAK-AAAAAEAAAKAAA 29

RESULT 10
ADE10634
ID   ADE10634 standard; protein; 104 AA.
XX
AC   ADE10634;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Structurally biased random peptide library related protein seqid 41.
XX
KW   fusion nucleic acid library; scaffold protein; bioactive peptide;
KW   phenotype change; cell morphology; cell growth; cell viability;
KW   cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW   loss of cell division; decreased cell growth; brca-1; brca-2;
KW   tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW   Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW   skin biology; cosmeceutical; endocrinology; infectious disease;
KW   drug toxicity; drug resistance; inflammation; allergic response.
XX
OS   Synthetic.
XX
PN   US2003143562-A1.
XX
PD   31-JUL-2003.
XX
PF   20-JUN-2002; 2002US-00177725.
XX
PR   08-OCT-1998;   98US-00169015.
PR   08-OCT-1999;   99US-00415765.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Anderson D,  Peelle BR,  Bogenberger JM;
XX
DR   WPI; 2003-829786/77.
XX
PT   Novel library of fusion nucleic acids each of which has fused first and
PT   second nucleic acids encoding scaffold protein and library peptide having
PT   alpha helical biasing sequence, respectively, useful in screening
PT   methods.
XX
PS   Example 6; SEQ ID NO 41; 110pp; English.
XX
CC   The invention describes a library (I) of fusion nucleic acids, where each
CC   fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC   scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC   library peptide sequence comprising an alpha helical biasing sequence;
CC   where N1 is fused to N2. Disclosed is a method for screening bioactive
CC   peptides conferring a change in specific phenotype such as cell
CC   morphology, cell growth, cell viability, adhesion to substrates or other
CC   cells, and cellular density; changes in the expression of one or more
CC   RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC   in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC   lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC   peptide identified by above mentioned method is used to generate more
CC   candidate peptides and to identify target molecules, i.e., the molecules
CC   with which the bioactive peptide interacts. The peptide(s) can be
CC   combined with other pharmacologic activators to study the epistatic
CC   relationships of signal transduction pathways in question. The disclosed
CC   method is also useful in cancer applications. Random libraries can be
CC   introduced into any tumour cell (primary or cultured), and peptides
CC   identified which by themselves induce apoptosis, cell death, loss of cell
CC   division or decreased cell growth. The method is also useful for
CC   screening of bioactive peptides which restore the constitutive function
CC   of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC   important in breast cancer such as the adenomatous polyposis coli gene
CC   (APC) and the Drosophila discs-large gene (DIg), which are components of
CC   cell-cell junctions. The methods are useful in cardiovascular
CC   applications, neurobiology applications, bone biology applications, skin
CC   biology applications, cosmeceutical applications, endocrinology
CC   applications, infectious disease applications, drug toxicities and drug
CC   resistance applications, immunobiology, inflammation, and allergic
CC   response applications, and biotechnology applications. The peptide
CC   library can easily be monitored, both for its presence within cells and
CC   its quantity. The expression of structurally biased libraries generate
CC   elevated cellular concentration of peptides having a given structural
CC   bias and thus increase the hit rate for targets that bind such
CC   structures. This is the amino acid sequence of a protein associated with
CC   fused nucleic acid and random peptide libraries of the invention.
XX
SQ   Sequence 104 AA;

  Query Match             58.6%;  Score 61.5;  DB 7;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.37;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
                | | ||| ||| || ||| |||| |
Db          6 AAAAAAEAAAK-AAAAAEAAAKAAA 29


RESULT 11
ADE10698
ID   ADE10698 standard; protein; 59 AA.
XX
AC   ADE10698;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Structurally biased random peptide library scaffold protein seqid 105.
XX
KW   fusion nucleic acid library; scaffold protein; bioactive peptide;
KW   phenotype change; cell morphology; cell growth; cell viability;
KW   cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW   loss of cell division; decreased cell growth; brca-1; brca-2;
KW   tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW   Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW   skin biology; cosmeceutical; endocrinology; infectious disease;
KW   drug toxicity; drug resistance; inflammation; allergic response;
KW   scaffold protein.
XX
OS   Synthetic.
XX
PN   US2003143562-A1.
XX
PD   31-JUL-2003.
XX
PF   20-JUN-2002; 2002US-00177725.
XX
PR   08-OCT-1998;   98US-00169015.
PR   08-OCT-1999;   99US-00415765.
XX
PA   (RIGE-) RIGEL PHARM INC.
XX
PI   Anderson D,  Peelle BR,  Bogenberger JM;
XX
DR   WPI; 2003-829786/77.
XX
PT   Novel library of fusion nucleic acids each of which has fused first and
PT   second nucleic acids encoding scaffold protein and library peptide having
PT   alpha helical biasing sequence, respectively, useful in screening
PT   methods.
XX
PS   Disclosure; SEQ ID NO 105; 110pp; English.
XX
CC   The invention describes a library (I) of fusion nucleic acids, where each
CC   fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC   scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC   library peptide sequence comprising an alpha helical biasing sequence;
CC   where N1 is fused to N2. Disclosed is a method for screening bioactive
CC   peptides conferring a change in specific phenotype such as cell
CC   morphology, cell growth, cell viability, adhesion to substrates or other
```

CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 59 AA;

Query Match 58.1%; Score 61; DB 7; Length 59;
Best Local Similarity 69.6%; Pred. No. 0.23;
Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              | ||| ||| ||||| |||   |
Db          5 AAAAEAAAKAAAEAAAKAAAEAA 27
```

RESULT 12
ADE10648
ID ADE10648 standard; protein; 59 AA.
XX
AC ADE10648;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library related protein seqid 55.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Example 6; SEQ ID NO 55; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a protein associated with
CC fused nucleic acid and random peptide libraries of the invention.
XX
SQ Sequence 59 AA;

Query Match 58.1%; Score 61; DB 7; Length 59;
Best Local Similarity 69.6%; Pred. No. 0.23;
Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              | ||| ||| ||||| |||   |
Db          5 AAAAEAAAKAAAEAAAKAAAEAA 27
```

RESULT 13
ADE10697
ID ADE10697 standard; protein; 67 AA.
XX
AC ADE10697;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library scaffold protein seqid 104.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;

KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response;
KW scaffold protein.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Disclosure; SEQ ID NO 104; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 67 AA;

Query Match 58.1%; Score 61; DB 7; Length 67;
Best Local Similarity 69.6%; Pred. No. 0.27;

Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 AEAAEKAAKYAAEAAEKAAKAXA 25
| ||| ||| ||||| ||| |
Db 5 AAAAEAAAKAAAEAAAKAAAEAA 27

RESULT 14
ADE10647
ID ADE10647 standard; protein; 67 AA.
XX
AC ADE10647;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library related protein seqid 54.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Example 6; SEQ ID NO 54; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene

CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a protein associated with
CC fused nucleic acid and random peptide libraries of the invention.
XX
SQ Sequence 67 AA;

Query Match 58.1%; Score 61; DB 7; Length 67;
Best Local Similarity 69.6%; Pred. No. 0.27;
Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              | ||| ||| ||||| |||   |
Db          5 AAAAEAAAKAAAEAAAKAAAEAA 27
```

RESULT 15
ADE10696
ID ADE10696 standard; protein; 75 AA.
XX
AC ADE10696;
XX
DT 29-JAN-2004 (first entry)
XX
DE Structurally biased random peptide library scaffold protein seqid 103.
XX
KW fusion nucleic acid library; scaffold protein; bioactive peptide;
KW phenotype change; cell morphology; cell growth; cell viability;
KW cell adhesion; cellular density; cancer; tumour; apoptosis; cell death;
KW loss of cell division; decreased cell growth; brca-1; brca-2;
KW tumour suppressor gene; breast cancer; adenomatous polyposis coli; APC;
KW Drosophila discs-large; DIg; cardiovascular; neurobiology; bone biology;
KW skin biology; cosmeceutical; endocrinology; infectious disease;
KW drug toxicity; drug resistance; inflammation; allergic response;
KW scaffold protein.
XX
OS Synthetic.
XX
PN US2003143562-A1.
XX
PD 31-JUL-2003.
XX
PF 20-JUN-2002; 2002US-00177725.
XX
PR 08-OCT-1998; 98US-00169015.
PR 08-OCT-1999; 99US-00415765.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Anderson D, Peelle BR, Bogenberger JM;
XX
DR WPI; 2003-829786/77.
XX
PT Novel library of fusion nucleic acids each of which has fused first and
PT second nucleic acids encoding scaffold protein and library peptide having
PT alpha helical biasing sequence, respectively, useful in screening
PT methods.
XX
PS Disclosure; SEQ ID NO 103; 110pp; English.
XX
CC The invention describes a library (I) of fusion nucleic acids, where each
CC fusion nucleic acid comprises a first nucleic acid (N1), encoding a
CC scaffold protein sequence; and a second nucleic acid (N2), encoding a
CC library peptide sequence comprising an alpha helical biasing sequence;
CC where N1 is fused to N2. Disclosed is a method for screening bioactive
CC peptides conferring a change in specific phenotype such as cell
CC morphology, cell growth, cell viability, adhesion to substrates or other
CC cells, and cellular density; changes in the expression of one or more
CC RNAs, proteins, lipids, hormones, cytokines, or other molecules; changes
CC in the equilibrium state (i.e., half-life) or one or more RNAs, protein,
CC lipids, hormones, cytokines, or other molecules; etc. The bioactive
CC peptide identified by above mentioned method is used to generate more
CC candidate peptides and to identify target molecules, i.e., the molecules
CC with which the bioactive peptide interacts. The peptide(s) can be
CC combined with other pharmacologic activators to study the epistatic
CC relationships of signal transduction pathways in question. The disclosed
CC method is also useful in cancer applications. Random libraries can be
CC introduced into any tumour cell (primary or cultured), and peptides
CC identified which by themselves induce apoptosis, cell death, loss of cell
CC division or decreased cell growth. The method is also useful for
CC screening of bioactive peptides which restore the constitutive function
CC of the brca-1 or brca-2 genes, and other tumour suppressor genes
CC important in breast cancer such as the adenomatous polyposis coli gene
CC (APC) and the Drosophila discs-large gene (DIg), which are components of
CC cell-cell junctions. The methods are useful in cardiovascular
CC applications, neurobiology applications, bone biology applications, skin
CC biology applications, cosmeceutical applications, endocrinology
CC applications, infectious disease applications, drug toxicities and drug
CC resistance applications, immunobiology, inflammation, and allergic
CC response applications, and biotechnology applications. The peptide
CC library can easily be monitored, both for its presence within cells and
CC its quantity. The expression of structurally biased libraries generate
CC elevated cellular concentration of peptides having a given structural
CC bias and thus increase the hit rate for targets that bind such
CC structures. This is the amino acid sequence of a scaffold protein used in
CC peptide libraries ot hold the library peptide in a conformationally
CC restricted form.
XX
SQ Sequence 75 AA;

Query Match 58.1%; Score 61; DB 7; Length 75;
Best Local Similarity 69.6%; Pred. No. 0.3;
Matches 16; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              | ||| ||| ||||| |||   |
Db          5 AAAAEAAAKAAAEAAAKAAAEAA 27
```

Search completed: April 20, 2004, 21:59:10
Job time : 56 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:58:10 ; Search time 23 Seconds
(without alignments)
56.115 Million cell updates/sec

Title: US-10-019-482-1
Perfect score: 105
Sequence: 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60	57.1	33	1	US-08-303-025-16	Sequence 16, Appl
2	60	57.1	33	2	US-08-436-703B-4	Sequence 4, Appli
3	59.5	56.7	56	4	US-09-405-743A-3	Sequence 3, Appli
4	59	56.2	469	4	US-09-489-039A-13565	Sequence 13565, A
5	57.5	54.8	67	4	US-09-869-875-7	Sequence 7, Appli
6	57.5	54.8	86	4	US-09-405-743A-6	Sequence 6, Appli
7	57.5	54.8	117	4	US-09-340-736E-9	Sequence 9, Appli
8	57.5	54.8	118	4	US-09-340-736E-10	Sequence 10, Appl
9	57.5	54.8	199	4	US-09-340-736E-11	Sequence 11, Appl
10	57.5	54.8	200	4	US-09-340-736E-2	Sequence 2, Appli
11	57.5	54.8	201	2	US-08-911-364-2	Sequence 2, Appli
12	57.5	54.8	731	2	US-08-911-364-1	Sequence 1, Appli
13	57.5	54.8	731	4	US-09-340-736E-1	Sequence 1, Appli
14	57.5	54.8	733	3	US-08-464-700-2	Sequence 2, Appli
15	56	53.3	45	4	US-09-405-743A-2	Sequence 2, Appli
16	56	53.3	92	4	US-09-344-529-2	Sequence 2, Appli
17	56	53.3	109	4	US-09-405-743A-7	Sequence 7, Appli
18	55	52.4	28	1	US-08-303-025-12	Sequence 12, Appl
19	55	52.4	28	2	US-08-436-703B-1	Sequence 1, Appli
20	55	52.4	29	1	US-08-152-488-10	Sequence 10, Appl
21	55	52.4	29	1	US-08-152-488-11	Sequence 11, Appl
22	55	52.4	29	1	US-08-152-488-12	Sequence 12, Appl
23	55	52.4	29	1	US-08-303-025-10	Sequence 10, Appl
24	55	52.4	29	1	US-08-303-025-11	Sequence 11, Appl
25	55	52.4	29	1	US-08-303-025-13	Sequence 13, Appl
26	55	52.4	29	1	US-08-303-025-14	Sequence 14, Appl
27	55	52.4	29	1	US-08-677-304-10	Sequence 10, Appl
28	55	52.4	29	1	US-08-677-304-11	Sequence 11, Appl
29	55	52.4	29	1	US-08-677-304-12	Sequence 12, Appl
30	55	52.4	29	2	US-08-436-703B-3	Sequence 3, Appli
31	55	52.4	29	2	US-08-436-703B-15	Sequence 15, Appl
32	55	52.4	29	2	US-08-436-703B-16	Sequence 16, Appl
33	55	52.4	32	1	US-08-152-488-13	Sequence 13, Appl
34	55	52.4	32	1	US-08-303-025-15	Sequence 15, Appl
35	55	52.4	32	1	US-08-677-304-13	Sequence 13, Appl
36	55	52.4	32	2	US-08-436-703B-2	Sequence 2, Appli
37	54	51.4	77	4	US-09-405-743A-5	Sequence 5, Appli
38	53.5	51.0	66	4	US-09-405-743A-4	Sequence 4, Appli
39	52	49.5	39	4	US-09-117-121-37	Sequence 37, Appl
40	52	49.5	39	4	US-09-117-121-38	Sequence 38, Appl
41	52	49.5	407	4	US-09-252-991A-29581	Sequence 29581, A
42	50.5	48.1	24	2	US-08-491-527A-13	Sequence 13, Appl
43	50.5	48.1	66	4	US-08-858-207A-312	Sequence 312, App
44	50	47.6	54	4	US-09-117-121-30	Sequence 30, Appl
45	49.5	47.1	39	4	US-09-117-121-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-303-025-16
; Sequence 16, Application US/08303025
; Patent No. 5614494
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR HEPARIN AND
; TITLE OF INVENTION: LOW MOLECULAR WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 150 West Jefferson, Suite 2500
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226-4415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy diskette 3.5" 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS v.6.22
; SOFTWARE: WordPerfect 6.1; ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/303,025
; FILING DATE: 08-SEPT-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06829
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 08/152,488
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REFERENCE/DOCKET NUMBER: 7WH-060548-00231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-496-7622
; TELEFAX: 313-496-8454
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A

; TITLE: N/A
; DOCUMENT NUMBER: PCT/US92/08069
; FILING DATE: 14-AUG-1993
US-08-303-025-16

Query Match 57.1%; Score 60; DB 1; Length 33;
Best Local Similarity 70.0%; Pred. No. 0.026;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 EAAEKAAKYAAEAAEKAAKA 23
||||:|||| | :||:|| ||
Db 1 EAAKKAAKKAKKAAKKAKKA 20

RESULT 2
US-08-436-703B-4
; Sequence 4, Application US/08436703B
; Patent No. 5919761
; GENERAL INFORMATION:
; APPLICANT: Wakefield, Thomas W.
; APPLICANT: Andrews, Philip C.
; APPLICANT: Stanley, James C.
; TITLE OF INVENTION: NOVEL PEPTIDES FOR
; TITLE OF INVENTION: HEPARIN AND LOW MOLECULAR
; TITLE OF INVENTION: WEIGHT HEPARIN
; TITLE OF INVENTION: ANTICOAGULATION REVERSAL
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benita J, Rohm, Esq.
; STREET: 6601 Woodward Avenue
; STREET: Suite 1525
; CITY: Detroit
; STATE: Michigan
; COUNTRY: United States of America
; ZIP: 48226
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk 1.44Mb, 3.5"
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6;
; SOFTWARE: ASCII (DOS)Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,703B
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: N/A
; FILING DATE: N/A
; ATTORNEY/AGENT INFORMATION:
; NAME: Rohm, Benita J.
; REGISTRATION NUMBER: 28,664
; REFERENCE/DOCKET NUMBER: 7WK-060548-00233
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 313-965-1976
; TELEFAX: 313-965-1951
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: N/A
; PUBLICATION INFORMATION:
; AUTHORS: N/A
; TITLE: N/A
US-08-436-703B-4

Query Match 57.1%; Score 60; DB 2; Length 33;
Best Local Similarity 70.0%; Pred. No. 0.026;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 EAAEKAAKYAAEAAEKAAKA 23
||||:|||| | :||:|| ||
Db 1 EAAKKAAKKAKKAAKKAKKA 20

RESULT 3
US-09-405-743A-3
; Sequence 3, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A
; CURRENT APPLICATION NUMBER: US/09/405,743A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-405-743A-3

Query Match 56.7%; Score 59.5; DB 4; Length 56;
Best Local Similarity 66.7%; Pred. No. 0.055;
Matches 16; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy 3 AEAAEKA-AKYAAEAAEKAAKAXA 25
||| :|| ||| |||||: ||| |
Db 30 AEAKKKAEAKYKAEAAKAAAKEAA 53

RESULT 4
US-09-489-039A-13565
; Sequence 13565, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13565
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13565

Query Match 56.2%; Score 59; DB 4; Length 469;
Best Local Similarity 65.2%; Pred. No. 0.7;
Matches 15; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 3 AEAAEKAAKYAAEAAEKAAKAXA 25
| |||||| ||:||||| |
Db 302 AAAAEKAAAAEKAAADKAAKAAA 324

RESULT 5
US-09-869-875-7
; Sequence 7, Application US/09869875
; Patent No. 6521456
; GENERAL INFORMATION:
; APPLICANT: Siebenkotten, Gregor
; APPLICANT: Christine, Rainer
; TITLE OF INVENTION: USE OF CELLULAR TRANSPORT SYSTEMS FOR THE TRANSFER OF NUCLEIC AC
; TITLE OF INVENTION: THROUGH THE NUCLEAR ENVELOPE

```
;  FILE REFERENCE: 30430.1USWO
;  CURRENT APPLICATION NUMBER: US/09/869,875
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: PCT/DE00/00061
;  PRIOR FILING DATE: 2000-01-03
;  PRIOR APPLICATION NUMBER: DE 199 00 513.3
;  PRIOR FILING DATE: 1999-01-08
;  PRIOR APPLICATION NUMBER: DE 199 33 939.2
;  PRIOR FILING DATE: 1999-07-20
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 67
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: PNA-NLS
US-09-869-875-7

  Query Match           54.8%;  Score 57.5;  DB 4;  Length 67;
  Best Local Similarity   61.5%;  Pred. No. 0.13;
  Matches   16;  Conservative    4;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          1 AXAEAAEKAAKYAA-EAAEKAAKAXA 25
              |  ||||:||: || ||||:||:  |
Db          4 AAEEAAEEAAEEAAEEAAEEAAEEAA 29


RESULT 6
US-09-405-743A-6
; Sequence 6, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 86
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-6

  Query Match           54.8%;  Score 57.5;  DB 4;  Length 86;
  Best Local Similarity   60.7%;  Pred. No. 0.17;
  Matches   17;  Conservative    1;  Mismatches    7;  Indels    3;  Gaps    1;

Qy          1 AXAEAAEKA---AKYAAEAAEKAAKAXA 25
              | ||  | |   ||| ||||:|| || |
Db         47 AKAEKKEYAAAEAKYKAEAAKKAYKAEA 74


RESULT 7
US-09-340-736E-9
; Sequence 9, Application US/09340736E
; Patent No. 6489446
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, ASER
;  APPLICANT:  KEELEY, FRED
;  APPLICANT:  ROTHSTEIN, STEVEN
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
;  TITLE OF INVENTION:  AND OTHER FIBROUS PROTEINS
;  FILE REFERENCE: 041082/0110
;  CURRENT APPLICATION NUMBER: US/09/340,736E
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER: 08/911,364
;  PRIOR FILING DATE: 1997-08-07
;  PRIOR APPLICATION NUMBER: 60/023,552
;  PRIOR FILING DATE: 1996-08-07
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 117
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: MFU-3 polypeptide
US-09-340-736E-9

  Query Match           54.8%;  Score 57.5;  DB 4;  Length 117;
  Best Local Similarity   60.7%;  Pred. No. 0.24;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db         37 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 64


RESULT 8
US-09-340-736E-10
; Sequence 10, Application US/09340736E
; Patent No. 6489446
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, ASER
;  APPLICANT:  KEELEY, FRED
;  APPLICANT:  ROTHSTEIN, STEVEN
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
;  TITLE OF INVENTION:  AND OTHER FIBROUS PROTEINS
;  FILE REFERENCE: 041082/0110
;  CURRENT APPLICATION NUMBER: US/09/340,736E
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER: 08/911,364
;  PRIOR FILING DATE: 1997-08-07
;  PRIOR APPLICATION NUMBER: 60/023,552
;  PRIOR FILING DATE: 1996-08-07
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 118
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: MFU-4 polypeptide
US-09-340-736E-10

  Query Match           54.8%;  Score 57.5;  DB 4;  Length 118;
  Best Local Similarity   60.7%;  Pred. No. 0.24;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db         38 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 65


RESULT 9
US-09-340-736E-11
; Sequence 11, Application US/09340736E
; Patent No. 6489446
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, ASER
;  APPLICANT:  KEELEY, FRED
;  APPLICANT:  ROTHSTEIN, STEVEN
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
;  TITLE OF INVENTION:  AND OTHER FIBROUS PROTEINS
;  FILE REFERENCE: 041082/0110
;  CURRENT APPLICATION NUMBER: US/09/340,736E
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER: 08/911,364
```

```
;  PRIOR FILING DATE: 1997-08-07
;  PRIOR APPLICATION NUMBER: 60/023,552
;  PRIOR FILING DATE: 1996-08-07
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 199
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: MFU-5 polypeptide
US-09-340-736E-11

  Query Match             54.8%;  Score 57.5;  DB 4;  Length 199;
  Best Local Similarity   60.7%;  Pred. No. 0.44;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db         37 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 64


RESULT 10
US-09-340-736E-2
; Sequence 2, Application US/09340736E
; Patent No. 6489446
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, ASER
;  APPLICANT:  KEELEY, FRED
;  APPLICANT:  ROTHSTEIN, STEVEN
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
;  TITLE OF INVENTION:  AND OTHER FIBROUS PROTEINS
;  FILE REFERENCE: 041082/0110
;  CURRENT APPLICATION NUMBER: US/09/340,736E
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER: 08/911,364
;  PRIOR FILING DATE: 1997-08-07
;  PRIOR APPLICATION NUMBER: 60/023,552
;  PRIOR FILING DATE: 1996-08-07
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: MFU-1 polypeptide
US-09-340-736E-2

  Query Match             54.8%;  Score 57.5;  DB 4;  Length 200;
  Best Local Similarity   60.7%;  Pred. No. 0.44;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db         38 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 65


RESULT 11
US-08-911-364-2
; Sequence 2, Application US/08911364
; Patent No. 5969106
;  GENERAL INFORMATION:
;    APPLICANT:  ROTHSTEIN, Aser
;    APPLICANT:  KEELY, Fred W.
;    APPLICANT:  ROTHSTEIN, Steven J.
;    TITLE OF INVENTION:  SELF-ALIGNING PEPTIDES MODELED ON HUMAN
;    TITLE OF INVENTION:  ELASTIN AND OTHER FIBROUS PROTEINS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/911,364
;      FILING DATE:  07-AUG-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/023,552
;      FILING DATE:  07-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  041082/0104
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  201 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-911-364-2

  Query Match             54.8%;  Score 57.5;  DB 2;  Length 201;
  Best Local Similarity   60.7%;  Pred. No. 0.44;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db         38 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 65


RESULT 12
US-08-911-364-1
; Sequence 1, Application US/08911364
; Patent No. 5969106
;  GENERAL INFORMATION:
;    APPLICANT:  ROTHSTEIN, Aser
;    APPLICANT:  KEELY, Fred W.
;    APPLICANT:  ROTHSTEIN, Steven J.
;    TITLE OF INVENTION:  SELF-ALIGNING PEPTIDES MODELED ON HUMAN
;    TITLE OF INVENTION:  ELASTIN AND OTHER FIBROUS PROTEINS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/911,364
;      FILING DATE:  07-AUG-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/023,552
;      FILING DATE:  07-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Bent, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  041082/0104
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 672-5300
;       TELEFAX:  (202) 672-5399
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  731 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-911-364-1

  Query Match             54.8%;  Score 57.5;  DB 2;  Length 731;
  Best Local Similarity   60.7%;  Pred. No. 1.9;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db        415 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 442


RESULT 13
US-09-340-736E-1
; Sequence 1, Application US/09340736E
; Patent No. 6489446
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, ASER
;  APPLICANT:  KEELEY, FRED
;  APPLICANT:  ROTHSTEIN, STEVEN
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
;  TITLE OF INVENTION:  AND OTHER FIBROUS PROTEINS
;  FILE REFERENCE: 041082/0110
;  CURRENT APPLICATION NUMBER: US/09/340,736E
;  CURRENT FILING DATE:  1999-06-29
;  PRIOR APPLICATION NUMBER: 08/911,364
;  PRIOR FILING DATE: 1997-08-07
;  PRIOR APPLICATION NUMBER: 60/023,552
;  PRIOR FILING DATE: 1996-08-07
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 731
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-340-736E-1

  Query Match             54.8%;  Score 57.5;  DB 4;  Length 731;
  Best Local Similarity   60.7%;  Pred. No. 1.9;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db        415 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 442


RESULT 14
US-08-464-700-2
; Sequence 2, Application US/08464700
; Patent No. 6232458
;  GENERAL INFORMATION:
;    APPLICANT:  WEISS, ANTHONY S
;    APPLICANT:  MARTIN, STEPHEN L
;    TITLE OF INVENTION:  SYNTHETIC POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,700
;      FILING DATE:  7-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PL6520
;      FILING DATE:  22-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PL9661
;      FILING DATE:  28-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU93/00655
;      FILING DATE:  16-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GHC3USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  733 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-464-700-2

  Query Match             54.8%;  Score 57.5;  DB 3;  Length 733;
  Best Local Similarity   60.7%;  Pred. No. 1.9;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db        417 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 444


RESULT 15
US-09-405-743A-2
; Sequence 2, Application US/09405743A
; Patent No. 6514938
; GENERAL INFORMATION:
;  APPLICANT:  Yeda Research and Development Co., Ltd.
;  TITLE OF INVENTION:  GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
;  FILE REFERENCE:  60807-A
;  CURRENT APPLICATION NUMBER: US/09/405,743A
;  CURRENT FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 45
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
;   OTHER INFORMATION: PEPTIDE
US-09-405-743A-2

  Query Match             53.3%;  Score 56;  DB 4;  Length 45;
  Best Local Similarity   60.9%;  Pred. No. 0.14;
  Matches   14;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              |  |:|||||   ||||||   |
Db         19 AAEAKKAAKYEKAAAEKAAAKEA 41
```

Search completed: April 20, 2004, 22:01:34
Job time : 23 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 22:00:30 ; Search time 42 Seconds
(without alignments)
164.091 Million cell updates/sec

Title: US-10-019-482-1
Perfect score: 105
Sequence: 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1124875 seqs, 275673149 residues

Total number of hits satisfying chosen parameters: 1124875

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	65	61.9	104	12	US-10-393-449-42	Sequence 42, Appl
2	65	61.9	104	12	US-10-393-449-92	Sequence 92, Appl
3	65	61.9	104	14	US-10-177-725-42	Sequence 42, Appl
4	65	61.9	104	14	US-10-177-725-92	Sequence 92, Appl
5	62.5	59.5	428	12	US-10-282-122A-55748	Sequence 55748, A
6	62	59.0	104	12	US-10-393-449-39	Sequence 39, Appl
7	62	59.0	104	12	US-10-393-449-40	Sequence 40, Appl
8	62	59.0	104	12	US-10-393-449-89	Sequence 89, Appl
9	62	59.0	104	12	US-10-393-449-90	Sequence 90, Appl
10	62	59.0	104	14	US-10-177-725-39	Sequence 39, Appl
11	62	59.0	104	14	US-10-177-725-40	Sequence 40, Appl
12	62	59.0	104	14	US-10-177-725-89	Sequence 89, Appl
13	62	59.0	104	14	US-10-177-725-90	Sequence 90, Appl
14	61.5	58.6	104	12	US-10-393-449-41	Sequence 41, Appl
15	61.5	58.6	104	12	US-10-393-449-91	Sequence 91, Appl
16	61.5	58.6	104	14	US-10-177-725-41	Sequence 41, Appl
17	61.5	58.6	104	14	US-10-177-725-91	Sequence 91, Appl
18	61	58.1	59	12	US-10-393-449-55	Sequence 55, Appl
19	61	58.1	59	12	US-10-393-449-105	Sequence 105, App
20	61	58.1	59	14	US-10-177-725-55	Sequence 55, Appl
21	61	58.1	59	14	US-10-177-725-105	Sequence 105, App
22	61	58.1	67	12	US-10-393-449-54	Sequence 54, Appl
23	61	58.1	67	12	US-10-393-449-104	Sequence 104, App
24	61	58.1	67	14	US-10-177-725-54	Sequence 54, Appl
25	61	58.1	67	14	US-10-177-725-104	Sequence 104, App
26	61	58.1	75	12	US-10-393-449-53	Sequence 53, Appl
27	61	58.1	75	12	US-10-393-449-103	Sequence 103, App
28	61	58.1	75	14	US-10-177-725-53	Sequence 53, Appl
29	61	58.1	75	14	US-10-177-725-103	Sequence 103, App
30	61	58.1	83	12	US-10-393-449-52	Sequence 52, Appl
31	61	58.1	83	12	US-10-393-449-102	Sequence 102, App
32	61	58.1	83	14	US-10-177-725-52	Sequence 52, Appl
33	61	58.1	83	14	US-10-177-725-102	Sequence 102, App
34	61	58.1	88	12	US-10-393-449-49	Sequence 49, Appl
35	61	58.1	88	12	US-10-393-449-99	Sequence 99, Appl
36	61	58.1	88	14	US-10-177-725-49	Sequence 49, Appl
37	61	58.1	88	14	US-10-177-725-99	Sequence 99, Appl
38	61	58.1	91	12	US-10-393-449-51	Sequence 51, Appl
39	61	58.1	91	12	US-10-393-449-101	Sequence 101, App
40	61	58.1	91	14	US-10-177-725-51	Sequence 51, Appl
41	61	58.1	91	14	US-10-177-725-101	Sequence 101, App
42	61	58.1	104	12	US-10-393-449-47	Sequence 47, Appl
43	61	58.1	104	12	US-10-393-449-97	Sequence 97, Appl
44	61	58.1	104	14	US-10-177-725-47	Sequence 47, Appl
45	61	58.1	104	14	US-10-177-725-97	Sequence 97, Appl

ALIGNMENTS

RESULT 1
US-10-393-449-42
; Sequence 42, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
; FILE REFERENCE: RIGL-007CIP3
; CURRENT APPLICATION NUMBER: US/10/393,449
; CURRENT FILING DATE: 2003-03-18
; PRIOR APPLICATION NUMBER: US 10/177,725
; PRIOR FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-393-449-42

Query Match 61.9%; Score 65; DB 12; Length 104;
Best Local Similarity 68.0%; Pred. No. 0.12;
Matches 17; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| ||||| |||   |
Db         10 AAAAAAEAAAKAAAEAAAKAAAEAA 34
```

RESULT 2

```
US-10-393-449-92
; Sequence 92, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 92
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-6
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 38-40, 42-44, 46-47, 49-51, 53-54, 56-58, 60-6
;   OTHER INFORMATION: 2, 64-65, and 67-69 can be any amino acid
US-10-393-449-92

  Query Match             61.9%;  Score 65;  DB 12;  Length 104;
  Best Local Similarity   68.0%;  Pred. No. 0.12;
  Matches   17;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| ||||| ||| |
Db         10 AAAAAAEAAAKAAAEAAAKAAAEAA 34


RESULT 3
US-10-177-725-42
; Sequence 42, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: A-66900-4/RMS/AMS
;  CURRENT APPLICATION NUMBER: US/10/177,725
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
US-10-177-725-42

  Query Match             61.9%;  Score 65;  DB 14;  Length 104;
  Best Local Similarity   68.0%;  Pred. No. 0.12;
  Matches   17;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| ||||| ||| |
Db         10 AAAAAAEAAAKAAAEAAAKAAAEAA 34


RESULT 4
US-10-177-725-92
; Sequence 92, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: A-66900-4/RMS/AMS
;  CURRENT APPLICATION NUMBER: US/10/177,725
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 92
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 38-40, 42-44, 46-47, 49-51, 53-54, 56-58, 60-
;   OTHER INFORMATION: 2, 64-65, and 67-69 can be any amino acid
US-10-177-725-92

  Query Match             61.9%;  Score 65;  DB 14;  Length 104;
  Best Local Similarity   68.0%;  Pred. No. 0.12;
  Matches   17;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| ||||| ||| |
Db         10 AAAAAAEAAAKAAAEAAAKAAAEAA 34


RESULT 5
US-10-282-122A-55748
; Sequence 55748, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55748
;   LENGTH: 428
;   TYPE: PRT
;   ORGANISM: Enterobacter cloacae
US-10-282-122A-55748

  Query Match             59.5%;  Score 62.5;  DB 12;  Length 428;
  Best Local Similarity   64.3%;  Pred. No. 1.2;
  Matches   18;  Conservative    2;  Mismatches    3;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAA-----KYAAEAAEKAAKA 23
              | ||||:|||     | |||||:||| |
Db        210 AEAEAAKKAAQEAEKKAAAEAAKKAAAA 237


RESULT 6
US-10-393-449-39
; Sequence 39, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
US-10-393-449-39

  Query Match             59.0%;  Score 62;  DB 12;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 7
US-10-393-449-40
; Sequence 40, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
US-10-393-449-40

  Query Match             59.0%;  Score 62;  DB 12;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 8
US-10-393-449-89
; Sequence 89, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 89
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-6
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
US-10-393-449-89
```

```
  Query Match             59.0%;  Score 62;  DB 12;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 9
US-10-393-449-90
; Sequence 90, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 90
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-6
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
US-10-393-449-90

  Query Match             59.0%;  Score 62;  DB 12;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 10
US-10-177-725-39
; Sequence 39, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: A-66900-4/RMS/AMS
;  CURRENT APPLICATION NUMBER: US/10/177,725
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
US-10-177-725-39

  Query Match             59.0%;  Score 62;  DB 14;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 11
US-10-177-725-40
; Sequence 40, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: A-66900-4/RMS/AMS
;  CURRENT APPLICATION NUMBER: US/10/177,725
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
US-10-177-725-40

  Query Match             59.0%;  Score 62;  DB 14;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 12
US-10-177-725-89
; Sequence 89, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
;  FILE REFERENCE: A-66900-4/RMS/AMS
;  CURRENT APPLICATION NUMBER: US/10/177,725
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 89
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
```

```
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-6
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
US-10-177-725-89

  Query Match             59.0%;  Score 62;  DB 14;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 13
US-10-177-725-90
; Sequence 90, Application US/10177725
; Publication No. US20030143562A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: A-66900-4/RMS/AMS
;  CURRENT APPLICATION NUMBER: US/10/177,725
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 90
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-6
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
US-10-177-725-90

  Query Match             59.0%;  Score 62;  DB 14;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.29;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||| |||  || ||| |||| |
Db          9 AAAEAAAKAA--AAAAAEAAAKAAA 31


RESULT 14
US-10-393-449-41
; Sequence 41, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT S
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
US-10-393-449-41

  Query Match             58.6%;  Score 61.5;  DB 12;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.34;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| || ||| |||| |
Db          6 AAAAAAEAAAK-AAAAAEAAAKAAA 29


RESULT 15
US-10-393-449-91
; Sequence 91, Application US/10393449
; Publication No. US20030224412A1
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David
;  APPLICANT:  Bogenberger, Jakob M.
;  APPLICANT:  Peele, Beau R.
;  TITLE OF INVENTION: STRUCTURALLY BIASED RANDOM PEPTIDE LIBRARIES BASED ON DIFFERENT
;  FILE REFERENCE: RIGL-007CIP3
;  CURRENT APPLICATION NUMBER: US/10/393,449
;  CURRENT FILING DATE:  2003-03-18
;  PRIOR APPLICATION NUMBER: US 10/177,725
;  PRIOR FILING DATE: 2002-06-20
;  PRIOR APPLICATION NUMBER: US 09/415,765
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: US 09/169,015
;  PRIOR FILING DATE: 1998-10-08
;  NUMBER OF SEQ ID NOS: 173
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 91
;   LENGTH: 104
;   TYPE: PRT
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (37)..(68)
;   OTHER INFORMATION: "Xaa" at positions 37-39, 41-43, 45-46, 48-50, 52-53, 55-57, 59-
;   OTHER INFORMATION: 1, 63-64 and 66-68 can be any amino acid
US-10-393-449-91

  Query Match             58.6%;  Score 61.5;  DB 12;  Length 104;
  Best Local Similarity   72.0%;  Pred. No. 0.34;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||| ||| || ||| |||| |
Db          6 AAAAAAEAAAK-AAAAAEAAAKAAA 29


Search completed: April 20, 2004, 22:07:40
Job time : 43 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:57:14 ; Search time 20 Seconds
(without alignments)
120.239 Million cell updates/sec

Title: US-10-019-482-1
Perfect score: 105
Sequence: 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	65.7	924	2	T06636	hypothetical prote
2	62	59.0	168	2	T34804	hypothetical prote
3	60	57.1	179	2	AF2908	50S ribosomal prot
4	60	57.1	179	2	F97683	50S ribosomal prot
5	59	56.2	421	2	JV0057	tolA protein - Esc
6	57	54.3	177	2	E87294	ATP synthase F0, B
7	56	53.3	354	1	GNVVSR	genome polyprotein
8	56	53.3	375	2	A71625	rifin PFB0035c - m
9	55	52.4	564	2	AH2328	ATP-binding protei
10	55	52.4	909	2	T06635	hypothetical prote
11	54	51.4	101	2	H59099	hypothetical prote
12	54	51.4	592	1	IKEBCA	colicin A - Citrob
13	53	50.5	97	2	S02376	antifreeze protein
14	53	50.5	394	2	F90725	membrane spanning
15	53	50.5	394	2	G85576	membrane spanning
16	53	50.5	555	2	S04909	embryonic protein
17	53	50.5	1110	2	I51116	NF-180 - sea lampr
18	53	50.5	1147	2	T35781	hypothetical prote
19	52	49.5	110	2	T37490	ribosomal protein
20	52	49.5	192	2	T26386	hypothetical prote
21	52	49.5	217	2	A26721	histone H1-gamma,
22	52	49.5	228	2	E87612	cytochrome c, memb
23	52	49.5	347	2	E83525	TolA protein PA097
24	52	49.5	356	2	A82152	tolA protein VC183
25	52	49.5	643	1	T07064	seed biotin-contai
26	52	49.5	1156	2	T34852	probable secreted
27	51.5	49.0	45	2	A05163	antifreeze protein
28	51.5	49.0	846	2	S52418	GTP-binding regula
29	51	48.6	40	1	FDFI8G	antifreeze protein
30	51	48.6	147	2	D86389	hypothetical prote
31	51	48.6	205	2	S19114	cgcr-1 protein - C
32	51	48.6	229	2	C43330	gene 7 protein - p
33	51	48.6	294	2	S32234	transcription anti
34	51	48.6	294	2	S41061	probable transcrip
35	51	48.6	388	2	AC0138	TolA colicin impor
36	51	48.6	4687	1	A39638	plectin - rat
37	50.5	48.1	1203	2	C95229	DNA-directed RNA p
38	50.5	48.1	1216	2	G98093	DNA-directed RNA p
39	50	47.6	104	1	H64327	conserved hypothet
40	50	47.6	250	2	T35875	hypothetical prote
41	50	47.6	288	2	S58219	ABA-inducible prot
42	50	47.6	310	2	T34809	ribosomal protein
43	50	47.6	376	2	AG0592	tolA protein [impo
44	50	47.6	474	1	PWQFB	H+-transporting tw
45	50	47.6	893	2	T38147	dolichyl-phosphate

ALIGNMENTS

RESULT 1
T06636
hypothetical protein T20K18.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06636
R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Me
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15790
A;Accession: T06636
A;Molecule type: DNA
A;Residues: 1-924 <BEV>
A;Cross-references: EMBL:AL049640; GSPDB:GN00062; ATSP:T20K18.130
A;Experimental source: cultivar Columbia; BAC clone T20K18
C;Genetics:
A;Gene: ATSP:T20K18.130
A;Map position: 4
A;Introns: 209/2; 699/3; 753/3; 785/2; 807/2; 853/3; 912/3

Query Match 65.7%; Score 69; DB 2; Length 924;
Best Local Similarity 68.0%; Pred. No. 0.54;
Matches 17; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | | :|||| |||| ||| || | |
Db        603 AAAGARDKAAKAAAEAREKAEKAAA 627
```

RESULT 2
T34804
hypothetical protein SC2E1.36 SC2E1.36 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T34804
R;Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z21557
A;Accession: T34804
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-168 <MUR>
A;Cross-references: EMBL:AL023797; PIDN:CAA19411.1; GSPDB:GN00070; SCOEDB:SC2E1.36
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC2E1.36

Query Match 59.0%; Score 62; DB 2; Length 168;
Best Local Similarity 65.2%; Pred. No. 0.88;
Matches 15; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 AXAEAAEKAAKYAAEAAEKAAKA 23
              | |:||||| : || ||:||| |
```

```
Db        106 AEAKAAEKAERAAARAAKKAAPA 128


RESULT 3
AF2908
50S ribosomal protein L19 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2908
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-179 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43684.1; PID:g17741210; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: rplS
A;Map position: circular chromosome
C;Superfamily: Escherichia coli ribosomal protein L19

  Query Match             57.1%;  Score 60;  DB 2;  Length 179;
  Best Local Similarity   72.0%;  Pred. No. 1.6;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAE--AAEKAAKA 23
              | | ||||||  |||  |||:||||
Db        149 AQALAAEKAAAEAAEAKAAEEAAKA 173


RESULT 4
F97683
50S ribosomal protein L19 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: F97683
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: F97683
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-179 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88423.1; PID:g15157917; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4900
A;Map position: circular chromosome
C;Superfamily: Escherichia coli ribosomal protein L19

  Query Match             57.1%;  Score 60;  DB 2;  Length 179;
  Best Local Similarity   72.0%;  Pred. No. 1.6;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAE--AAEKAAKA 23
              | | ||||||  |||  |||:||||
Db        149 AQALAAEKAAAEAAEAKAAEEAAKA 173


RESULT 5
JV0057
tolA protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 01-Mar-2002
C;Accession: JV0057; B64810
R;Levengood, S.K.; Webster, R.E.
J. Bacteriol. 171, 6600-6609, 1989
A;Title: Nucleotide sequences of the tolA and tolB genes and localization of their produ
A;Reference number: JV0057; MUID:90078104; PMID:2687247
A;Accession: JV0057
A;Molecule type: DNA
A;Residues: 1-421 <LEV>
A;Cross-references: GB:M28232; NID:g148018; PIDN:AAA24683.1; PID:g148019
A;Experimental source: strain JM105
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64810
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-421 <BLAT>
A;Cross-references: GB:AE000177; GB:U00096; NID:g1786955; PIDN:AAC73833.1; PID:g1786960,
A;Experimental source: strain K-12, substrain MG1655
C;Comment: tolA and tolB proteins are necessary for colicins E2, E3, A, and K to reach r
C;Genetics:
A;Gene: tolA
A;Map position: 17 min
A;Start codon: GTG
C;Keywords: nucleotide binding; P-loop; transmembrane protein
F;14-34/Domain: transmembrane #status predicted <MSS>
F;78-301/Domain: helical #status predicted <HSR>
F;355-362/Region: nucleotide-binding motif A (P-loop)

  Query Match             56.2%;  Score 59;  DB 2;  Length 421;
  Best Local Similarity   60.0%;  Pred. No. 4.4;
  Matches   15;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |:|||:||| ||  |:| |:| |
Db        151 ADAKAAEEAAKKAAADAKKKAEAEA 175


RESULT 6
E87294
ATP synthase F0, B' subunit [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: E87294
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87294
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-177 <STO>
A;Cross-references: GB:AE005673; NID:g13421521; PIDN:AAK22353.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0366

  Query Match             54.3%;  Score 57;  DB 2;  Length 177;
  Best Local Similarity   60.0%;  Pred. No. 3.7;
  Matches   15;  Conservative    1;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||||: ||  |  ||| | | |
Db        110 ASAEAAERQAKEEAVLAEKLAAAEA 134


RESULT 7
GNVVSR
genome polyprotein 1 - tomato ringspot virus (strain raspberry) (fragment)
C;Species: tomato ringspot virus
```

C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999
C;Accession: A40787
R;Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
Virology 185, 468-472, 1991
A;Title: Comparison of the 5' and 3' termini of tomato ringspot virus RNA1 and RNA2: evi
A;Reference number: A40787; MUID:92024112; PMID:1926788
A;Accession: A40787
A;Molecule type: genomic RNA
A;Residues: 1-354 <ROT>
A;Cross-references: GB:M73822; NID:g335267; PIDN:AAA47941.1; PID:g555406
C;Genetics:
A;Map position: segment 1
C;Superfamily: tomato ringspot virus genome polyprotein
C;Keywords: glycoprotein; polyprotein
F;270/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.3%; Score 56; DB 1; Length 354;
Best Local Similarity 70.0%; Pred. No. 8.7;
Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 6 AEKAAKYAAEAAEKAAKAXA 25
| ||||||| || | | | |
Db 180 ARKAAKYAAFAARKKAAAVA 199

RESULT 8
A71625
rifin PFB0035c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 02-Mar-2001
C;Accession: A71625
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71625
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-375 <GAR>
A;Cross-references: GB:AE001367; GB:AE001362; NID:g3845074; PIDN:AAC71797.1; PID:g384507
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0035c
C;Superfamily: Plasmodium falciparum rifin PFB1005w

Query Match 53.3%; Score 56; DB 2; Length 375;
Best Local Similarity 65.0%; Pred. No. 9.1;
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 4 EAAEKAAKYAAEAAEKAAKA 23
| ||:||| ||:|||| |
Db 294 EGAEQAAKAAAKAAEKGVTA 313

RESULT 9
AH2328
ATP-binding protein of ABC transporter all4183 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AH2328
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2328
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-564 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75882.1; PID:g17133318; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all4183
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 52.4%; Score 55; DB 2; Length 564;
Best Local Similarity 65.0%; Pred. No. 17;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AEAAEKAAKYAAEAAEKAAK 22
| ||||||| ||:|: :||
Db 543 AIAAEKAAKKAAKSAKSSAK 562

RESULT 10
T06635
hypothetical protein T20K18.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06635
R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Me
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15790
A;Accession: T06635
A;Molecule type: DNA
A;Residues: 1-909 <BEV>
A;Cross-references: EMBL:AL049640; GSPDB:GN00062; ATSP:T20K18.120
A;Experimental source: cultivar Columbia; BAC clone T20K18
C;Genetics:
A;Gene: ATSP:T20K18.120
A;Map position: 4
A;Introns: 205/2; 686/3; 740/3; 772/2; 808/3; 838/3; 897/3

Query Match 52.4%; Score 55; DB 2; Length 909;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 14; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AXAEAAEKAAKYAAEAAEKAA 21
| ||| |:|| ||| |||||
Db 593 AHAEARERAAAGAREKAEKAA 613

RESULT 11
H59099
hypothetical protein pXO1-72 - Bacillus anthracis virulence plasmid pXO1
C;Species: Bacillus anthracis
C;Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C;Accession: H59099
R;Okinaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehle
J. Bacteriol. 181, 6509-6515, 1999
A;Title: Sequence and organization of pXO1, the large Bacillus anthracis plasmid harbor
A;Reference number: A59091; MUID:99445483; PMID:10515943
A;Accession: H59099
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-101 <OKI>
A;Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32376.1; PID:g4894288
A;Experimental source: strain Sterne
A;Note: similar to hypothetical, locus Clo tetP Clostridium perfringens (L20800)
C;Genetics:
A;Gene: pXO1-72
A;Genome: plasmid
C;Superfamily: Bacillus anthracis virulence plasmid pXO1 hypothetical protein pXO1-72

Query Match 51.4%; Score 54; DB 2; Length 101;
Best Local Similarity 63.6%; Pred. No. 5.2;
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AXAEAAEKAAKYAAEAAEKAAK 22
| :||||||| ||| | | |
Db 44 AEEKAAEKAAKEAAEEARIATK 65

RESULT 12

```
IKEBCA
colicin A - Citrobacter freundii (strain CA31) plasmid ColA
C;Species: Citrobacter freundii
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999
C;Accession: I40784; A03504; I40777
R;Morlon, J.; Chartier, M.; Bidaud, M.; Lazdunski, C.
Mol. Gen. Genet. 211, 231-243, 1988
A;Title: The complete nucleotide sequence of the colicinogenic plasmid ColA. High extent
A;Reference number: I40778; MUID:88174422; PMID:2832701
A;Accession: I40784
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-592 <RES>
A;Cross-references: GB:M37402; NID:g144661; PIDN:AAA72879.1; PID:g144667
A;Experimental source: plasmid ColA
R;Morlon, J.; Lloubes, R.; Varenne, S.; Chartier, M.; Lazdunski, C.
J. Mol. Biol. 170, 271-285, 1983
A;Title: Complete nucleotide sequence of the structural gene for colicin A, a gene trans
A;Reference number: A03504; MUID:84036205; PMID:6313941
A;Accession: A03504
A;Molecule type: DNA
A;Residues: 1-592 <MOR>
A;Cross-references: GB:X01008; GB:X00034; NID:g40459; PIDN:CAA25503.1; PID:g40460
R;Morlon, J.; Lloubes, R.; Chartier, M.; Bonicel, J.; Lazdunski, C.
EMBO J. 2, 787-789, 1983
A;Title: Nucleotide sequence of promoter, operator and amino-terminal region of caa, the
A;Reference number: I40777; MUID:84057757; PMID:6641715
A;Accession: I40777
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-53,'X',55-70 <RE2>
A;Cross-references: GB:M26369; NID:g144659; PIDN:AAA98057.1; PID:g144660
A;Experimental source: plasmid ColA
C;Comment: This protein acts to depolarize the bacterial inner membrane, most likely by
C;Genetics:
A;Gene: caa
A;Genome: plasmid
C;Superfamily: colicin IB
C;Keywords: antibiotic; bacteriocin; toxin; transmembrane protein

  Query Match             51.4%;  Score 54;  DB 1;  Length 592;
  Best Local Similarity   56.5%;  Pred. No. 23;
  Matches   13;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKA 23
              | ||||||| :  || ||:  :|
Db        364 AAAEAAEKARQRQAEEAERQRQA 386


RESULT 13
S02376
antifreeze protein precursor - yellowtail flounder
C;Species: Limanda ferruginea (yellowtail flounder)
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 24-Oct-2000
C;Accession: S02376
R;Scott, G.K.; Davies, P.L.; Shears, M.A.; Fletcher, G.L.
Eur. J. Biochem. 168, 629-633, 1987
A;Title: Structural variations in the alanine-rich antifreeze proteins of the Pleuronect
A;Reference number: S02376; MUID:88029483; PMID:3665937
A;Accession: S02376
A;Molecule type: mRNA
A;Residues: 1-97 <SCO>
A;Cross-references: EMBL:X06356; NID:g64041; PIDN:CAA29655.1; PID:g64042
A;Note: part of this sequence, including the amino end of the mature protein, was confir
C;Superfamily: antifreeze protein
C;Keywords: antifreeze
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-48/Domain: propeptide #status predicted <PRO>
F;49-96/Product: antifreeze protein #status predicted <MAT>

  Query Match             50.5%;  Score 53;  DB 2;  Length 97;
  Best Local Similarity   56.0%;  Pred. No. 6.6;
  Matches   14;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |  |  ||| ||: |  |||| |
Db         57 AAAATAAAAAKAAADTAAAAAKAAA 81


RESULT 14
F90725
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90725
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90725
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34197.1; PID:g13360233; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0774

  Query Match             50.5%;  Score 53;  DB 2;  Length 394;
  Best Local Similarity   56.0%;  Pred. No. 22;
  Matches   14;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              |  :|||:||| ||  |:| |:| |
Db        151 ADDKAAEEAAKKAAADAKKKAEAEA 175


RESULT 15
G85576
membrane spanning protein TolA [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Nov-2001
C;Accession: G85576
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85576
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <STO>
A;Cross-references: GB:AE005174; NID:g12513672; PIDN:AAG55075.1; GSPDB:GN00145; UWGP:Z0
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: tolA

  Query Match             50.5%;  Score 53;  DB 2;  Length 394;
  Best Local Similarity   56.0%;  Pred. No. 22;
  Matches   14;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              |  :|||:||| ||  |:| |:| |
Db        151 ADDKAAEEAAKKAAADAKKKAEAEA 175


Search completed: April 20, 2004, 22:01:00
Job time : 22 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:53:29 ; Search time 11 Seconds
(without alignments)
118.341 Million cell updates/sec

Title: US-10-019-482-1
Perfect score: 105
Sequence: 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60	57.1	179	1	RL19_AGRT5	Q8ubz5 agrobacteri
2	59	56.2	421	1	TOLA_ECOLI	P19934 escherichia
3	57.5	54.8	730	1	ELS_HUMAN	P15502 homo sapien
4	56	53.3	354	1	POL1_TRSVR	P29150 tomato ring
5	54.5	51.9	668	1	FAU_DROME	Q9vgx3 drosophila
6	54	51.4	592	1	CEA_CITFR	P04480 citrobacter
7	54	51.4	707	1	HS88_NEUCR	O74225 neurospora
8	53	50.5	97	1	ANP_LIMFE	P09031 limanda fer
9	53	50.5	177	1	RL19_RHIME	Q92139 rhizobium m
10	53	50.5	555	1	LED8_DAUCA	P20075 daucus caro
11	52	49.5	217	1	H1G_STRPU	P07796 strongyloce
12	52	49.5	347	1	TOLA_PSEAE	P50600 pseudomonas
13	51.5	49.0	45	1	ANP8_MYOSC	P04368 myoxocephal
14	51.5	49.0	495	1	AB31_CHLRE	Q8s339 chlamydomon
15	51.5	49.0	556	1	PT1_STRCO	Q9kzp1 streptomyce
16	51	48.6	40	1	ANP8_MYOAE	P20617 myoxocephal
17	51	48.6	229	1	VG07_BPP22	Q01074 bacteriopha
18	51	48.6	294	1	NUSG_STRGR	P36260 streptomyce
19	51	48.6	4473	1	PLE1_CRIGR	Q9ji55 cricetulus
20	51	48.6	4687	1	PLE1_RAT	P30427 rattus norv
21	50	47.6	104	1	Y223_METJA	Q57676 methanococc
22	50	47.6	168	1	RS16_COREF	Q8fp30 corynebacte
23	50	47.6	310	1	RS2_STRCO	O31212 streptomyce
24	50	47.6	474	1	ATPB_RHORU	P05038 rhodospiril
25	50	47.6	518	1	TPM4_DROME	P49455 drosophila
26	50	47.6	893	1	PMTX_SCHPO	O13898 schizosacch
27	50	47.6	902	1	IF2_BRAJA	Q89wa9 bradyrhizob
28	50	47.6	1882	1	POL2_TRSVR	P25247 tomato ring
29	49.5	47.1	181	1	RL19_RHILO	P58168 rhizobium l
30	49	46.7	156	1	H2B2_CHLRE	P54345 chlamydomon
31	49	46.7	248	1	H1_PARAN	P02256 parechinus
32	49	46.7	962	1	IF2_NEIMA	Q9jtb5 neisseria m
33	49	46.7	962	1	IF2_NEIMB	Q9jyd2 neisseria m
34	48.5	46.2	184	1	RS16_BACTN	Q9rq15 bacteroides
35	48.5	46.2	210	1	H1_LYTPI	P06144 lytechinus
36	48	45.7	124	1	RS16_RHIME	Q92143 rhizobium m
37	48	45.7	134	1	RS16_BRUME	Q8yj59 brucella me
38	48	45.7	300	1	NUSG_STRCO	P36266 streptomyce
39	48	45.7	372	1	TOLA_HAEIN	P44678 haemophilus
40	48	45.7	384	1	TMPB_TREPH	P29720 treponema p
41	48	45.7	433	1	ZUO1_YEAST	P32527 saccharomyc
42	48	45.7	907	1	IF2_VIBVU	Q8dbw0 vibrio vuln
43	48	45.7	907	1	IF2_VIBVY	Q7mi09 vibrio vuln
44	48	45.7	1130	1	YL17_CAEEL	Q11102 caenorhabdi
45	47.5	45.2	1009	1	IF2_CAUCR	Q9ac25 caulobacter

ALIGNMENTS

RESULT 1
RL19_AGRT5
ID RL19_AGRT5 STANDARD; PRT; 179 AA.
AC Q8UBZ5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L19.
GN RPLS OR ATU2703 OR AGR_C_4900.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
CC -!- FUNCTION: This protein is located at the 30S-50S ribosomal subunit
CC interface and may play a role in the structure and function of the
CC aminoacyl-tRNA binding site (By similarity).
CC -!- SIMILARITY: Belongs to the L19P family of ribosomal proteins.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AE009216; AAL43684.1; -.
DR EMBL; AE008183; AAK88423.1; -.

```
DR   PIR; AF2908; AF2908.
DR   PIR; F97683; F97683.
DR   HAMAP; MF_00402; -; 1.
DR   InterPro; IPR001857; Ribosomal_L19.
DR   Pfam; PF01245; Ribosomal_L19; 1.
DR   PRINTS; PR00061; RIBOSOMALL19.
DR   ProDom; PD002979; Ribosomal_L19; 1.
DR   TIGRFAMs; TIGR01024; rplS_bact; 1.
DR   PROSITE; PS01015; RIBOSOMAL_L19; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   179 AA;  19474 MW;  F3256BA44A5AD2D1 CRC64;

  Query Match             57.1%;  Score 60;  DB 1;  Length 179;
  Best Local Similarity   72.0%;  Pred. No. 0.65;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    2;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAE--AAEKAAKA 23
              | | ||||||  |||  |||:||||
Db        149 AQALAAEKAAAEAAEAKAAEEAAKA 173


RESULT 2
TOLA_ECOLI
ID   TOLA_ECOLI     STANDARD;      PRT;   421 AA.
AC   P19934;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   TolA protein.
GN   TOLA OR CIM OR EXCC OR LKY OR B0739.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JM105;
RX   MEDLINE=90078104; PubMed=2687247;
RA   Levengood S.K., Webster R.E.;
RT   "Nucleotide sequences of the tolA and tolB genes and localization of
RT   their products, components of a multistep translocation system in
RT   Escherichia coli.";
RL   J. Bacteriol. 171:6600-6609(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   DOMAINS.
RX   MEDLINE=91296736; PubMed=2068069;
RA   Levengood S.K., Beyer W.F. Jr., Webster R.E.;
RT   "TolA: a membrane protein involved in colicin uptake contains an
RT   extended helical region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN   [5]
RP   INTERACTION WITH PORINS.
RX   MEDLINE=97133271; PubMed=8978668;
RA   Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,
RA   Lloubes R.;
RT   "TolA central domain interacts with Escherichia coli porins.";
RL   EMBO J. 15:6408-6415(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX   MEDLINE=99332679; PubMed=10404600;
RA   Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;
RT   "Filamentous phage infection: crystal structure of g3p in complex
RT   with its coreceptor, the C-terminal domain of TolA.";
RL   Structure 7:711-722(1999).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
CC       COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC       COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC       BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
CC       OF BACTERIOPHAGE DNA.
CC   -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
CC       AND LAMB.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28232; AAA24683.1; -.
DR   EMBL; AE000177; AAC73833.1; -.
DR   EMBL; D90713; BAA35405.1; -.
DR   PIR; JV0057; JV0057.
DR   PDB; 1TOL; 20-MAY-99.
DR   EcoGene; EG11007; tolA.
KW   Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW   Repeat; Inner membrane; 3D-structure; Complete proteome.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35    421       PERIPLASMIC (POTENTIAL).
FT   DOMAIN       48    310       DOMAIN II (ALPHA-HELICAL).
FT   DOMAIN      311    421       DOMAIN III (FUNCTIONAL).
FT   DOMAIN      224    278       10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT                                A(2,4).
FT   DISULFID    363    388
FT   HELIX       335    349
FT   TURN        350    351
FT   TURN        353    354
FT   HELIX       355    358
FT   TURN        359    360
FT   STRAND      363    369
FT   TURN        371    372
FT   STRAND      375    383
FT   HELIX       385    397
FT   HELIX       406    412
FT   TURN        413    414
FT   STRAND      416    421
SQ   SEQUENCE   421 AA;  43156 MW;  8B2F52B4B97C655E CRC64;

  Query Match             56.2%;  Score 59;  DB 1;  Length 421;
  Best Local Similarity   60.0%;  Pred. No. 1.7;
  Matches   15;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |:|||:||| ||  |:| |:| |
Db        151 ADAKAAEEAAKKAAADAKKKAEAEA 175


RESULT 3
ELS_HUMAN
ID   ELS_HUMAN      STANDARD;      PRT;   730 AA.
```

```
AC   P15502; Q14233; Q14238;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Elastin precursor (Tropoelastin).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM B).
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89009960; PubMed=3171221;
RA   Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT   "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library: further elucidation of alternative splicing
RT   utilizing exon-specific oligonucleotides.";
RL   J. Invest. Dermatol. 91:458-464(1988).
RN   [3]
RP   SEQUENCE OF 164-724 FROM N.A. (ISOFORM B).
RC   TISSUE=Placenta;
RX   MEDLINE=88156138; PubMed=2831431;
RA   Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
RA   Rosenbloom J., Uitto J.;
RT   "Isolation and characterization of human elastin cDNAs, and age-
RT   associated variation in elastin gene expression in cultured skin
RT   fibroblasts.";
RL   Lab. Invest. 58:270-277(1988).
RN   [4]
RP   SEQUENCE OF 603-730 FROM N.A.
RC   TISSUE=Hippocampus, and Placenta;
RX   MEDLINE=96291399; PubMed=8689688;
RA   Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA   Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA   Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA   Odelberg S.J., Keating M.T.;
RT   "LIM-kinase1 hemizygosity implicated in impaired visuospatial
RT   constructive cognition.";
RL   Cell 86:59-69(1996).
CC   -!- FUNCTION: Major structural protein of tissues such as aorta and
CC       nuchal ligament, which must expand rapidly and recover completely.
CC   -!- SUBUNIT: The polymeric elastin chains are cross-linked together
CC       into an extensible 3D network.
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=P15502-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P15502-2; Sequence=VSP_004243;
CC   -!- PTM: The crosslinks are made of deaminated Lys.
CC   -!- DISEASE: Haploinsufficiency of ELN may be the cause of certain
CC       cardiovascular and musculo-skeletal abnormalities observed in
CC       Williams-Beuren syndrome (WBS), a rare developmental disorder. It
CC       is a contiguous gene deletion syndrome involving genes from
CC       chromosome band 7q11.23.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M17282; AAC98394.1; -.
DR   EMBL; M16983; AAC98394.1; JOINED.
DR   EMBL; M17265; AAC98394.1; JOINED.
DR   EMBL; M17266; AAC98394.1; JOINED.
DR   EMBL; M17267; AAC98394.1; JOINED.
DR   EMBL; M17268; AAC98394.1; JOINED.
DR   EMBL; M17270; AAC98394.1; JOINED.
DR   EMBL; M17271; AAC98394.1; JOINED.
DR   EMBL; M17272; AAC98394.1; JOINED.
DR   EMBL; M17273; AAC98394.1; JOINED.
DR   EMBL; M17275; AAC98394.1; JOINED.
DR   EMBL; M17276; AAC98394.1; JOINED.
DR   EMBL; M17277; AAC98394.1; JOINED.
DR   EMBL; M17278; AAC98394.1; JOINED.
DR   EMBL; M17279; AAC98394.1; JOINED.
DR   EMBL; M17280; AAC98394.1; JOINED.
DR   EMBL; M17281; AAC98394.1; JOINED.
DR   EMBL; M36860; AAA52382.1; -.
DR   EMBL; M24782; AAA53190.1; -.
DR   EMBL; U62292; AAB17544.1; -.
DR   EMBL; X15603; CAA33627.1; -.
DR   PIR; A32707; EAHU.
DR   HSSP; P50099; 1ZFJ.
DR   Genew; HGNC:3327; ELN.
DR   MIM; 130160; -.
DR   MIM; 194050; -.
DR   GO; GO:0005578; C:extracellular matrix; TAS.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR   GO; GO:0008283; P:cell proliferation; TAS.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   GO; GO:0007585; P:respiratory gaseous exchange; TAS.
DR   InterPro; IPR003979; tropoelastin.
DR   PRINTS; PR01500; TROPOELASTIN.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Williams-Beuren syndrome; Alternative splicing.
FT   SIGNAL       1     26
FT   CHAIN       27    730       ELASTIN.
FT   DISULFID   720    725       BY SIMILARITY.
FT   VARSPLIC   472    477       Missing (in isoform 2).
FT                               /FTId=VSP_004243.
SQ   SEQUENCE   730 AA;  63260 MW;  AB06D15BA567AE46 CRC64;

  Query Match             54.8%;  Score 57.5;  DB 1;  Length 730;
  Best Local Similarity   60.7%;  Pred. No. 4.1;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db        441 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 468


RESULT 4
POL1_TRSVR
ID   POL1_TRSVR     STANDARD;      PRT;   354 AA.
AC   P29150; Q88875;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RNA1 polyprotein (Fragment).
OS   Tomato ringspot virus (isolate raspberry) (TomRSV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC   Nepovirus.
OX   NCBI_TaxID=12281;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92024112; PubMed=1926788;
RA   Rott M.E., Tremaine J.H., Rochon D.M.;
RT   "Comparison of the 5' and 3' termini of tomato ringspot virus RNA1
```

```
RT   and RNA2: evidence for RNA recombination.";
RL   Virology 185:468-472(1991).
CC   -!- SIMILARITY: IDENTICAL FOR THE FIRST 132 AA, AND 75.3% IDENTICAL
CC       FOR THE NEXT 145 AA TO THE RNA2 POLYPROTEIN.
CC   -!- CAUTION: It is uncertain whether Met-1 or Met-122 is the
CC       initiator.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73822; AAA47941.1; -.
DR   EMBL; M73822; AAA47942.1; ALT_INIT.
DR   PIR; A40787; GNVVSR.
KW   Polyprotein; Coat protein.
FT   NON_TER     354    354
SQ   SEQUENCE   354 AA;  38338 MW;  7A26DE8258A3360B CRC64;

  Query Match             53.3%;  Score 56;  DB 1;  Length 354;
  Best Local Similarity   70.0%;  Pred. No. 3.5;
  Matches   14;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          6 AEKAAKYAAEAAEKAAKAXA 25
              | ||||||| || | | | |
Db        180 ARKAAKYAAFAARKKAAAVA 199


RESULT 5
FAU_DROME
ID   FAU_DROME      STANDARD;      PRT;   668 AA.
AC   Q9VGX3; Q95S18; Q9VGX1; Q9VGX2; Q9Y0F9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Anoxia upregulated protein.
GN   FAU OR CG6544.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S; TISSUE=Head;
RX   MEDLINE=99097004; PubMed=9878744;
RA   Ma E., Xu T., Haddad G.G.;
RT   "Gene regulation by O2 deprivation: an anoxia-regulated novel gene in
RT   Drosophila melanogaster.";
RL   Brain Res. Mol. Brain Res. 63:217-224(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM E).
RC   STRAIN=Berkeley; TISSUE=Head;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: Plays an important role in the regulation of tissue
CC       responsiveness to oxygen deprivation.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=A;
CC         IsoId=Q9VGX3-1; Sequence=Displayed;
CC       Name=B;
CC         IsoId=Q9VGX3-2; Sequence=VSP_004048, VSP_004049;
CC       Name=C;
CC         IsoId=Q9VGX3-3; Sequence=VSP_004046, VSP_004047;
CC       Name=D;
CC         IsoId=Q9VGX3-4; Sequence=VSP_004050, VSP_004051;
CC       Name=E;
CC         IsoId=Q9VGX3-5; Sequence=VSP_004052;
CC   -!- TISSUE SPECIFICITY: Concentrated in lamina neurons, first optic
CC       lobe neurons and cortical neurons of central brain.
CC   -!- INDUCTION: By anoxia.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF154418; AAD38397.1; -.
```

```
DR   EMBL; AE003688; AAF54549.2; -.
DR   EMBL; AE003688; AAF54550.1; -.
DR   EMBL; AE003688; AAF54551.1; -.
DR   EMBL; AE003688; AAF54552.1; -.
DR   EMBL; AY060997; AAL28545.1; -.
DR   EMBL; AY119569; AAM50223.1; -.
DR   FlyBase; FBgn0020439; fau.
DR   GO; GO:0006979; P:response to oxidative stress; IDA.
KW   Alternative splicing.
FT   DOMAIN       88    143       THR-RICH.
FT   DOMAIN      354    668       ALA/GLU-RICH.
FT   VARSPLIC     26    131       YPSVEKVTRVYKSSYPIYSSYSVPRRVYGATRVVTSPIRVV
FT                                TSPARVVSRVIHSPSPVRVVRTTTRVISSPERTTYSYTTPS
FT                                TYYSPSYLPSTYTSTYIPTSYTTY -> TPRLDLCTDRPGS
FT                                HRSRASSDYSYTSKSSVEKSSYDSSNPHSYRPERSTYTSTV
FT                                EKTSRSGPGGSYNYSTERTSTTGAGPGGYSYSSTTSGNLPG
FT                                GTKYRHFSYHV (in isoform C).
FT                                /FTId=VSP_004046.
FT   VARSPLIC    132    668       Missing (in isoform C).
FT                                /FTId=VSP_004047.
FT   VARSPLIC     26    163       YPSVEKVTRVYKSSYPIYSSYSVPRRVYGATRVVTSPIRVV
FT                                TSPARVVSRVIHSPSPVRVVRTTTRVISSPERTTYSYTTPS
FT                                TYYSPSYLPSTYTSTYIPTSYTTYTPSYAYSPTTVTRVYAP
FT                                RSSLSPLRITPSPVR -> RTKRTPIDWEKVPFVPRPSLIS
FT                                DPVTAFGVRRPDLERRQRSILDPINRASIKPDYKLAYEPIE
FT                                PYVSTRDKNRTRILGMVRQHIDTVEAGGNTAGRTFRDSLDA
FT                                QLPRLHRAVSESLPVRRETYRNERSGAMVTKYSY (in
FT                                isoform B).
FT                                /FTId=VSP_004048.
FT   VARSPLIC    164    668       Missing (in isoform B).
FT                                /FTId=VSP_004049.
FT   VARSPLIC     36     98       YKSSYPIYSSYSVPRRVYGATRVVTSPIRVVTSPARVVSRV
FT                                IHSPSPVRVVRTTTRVISSPER -> SIWFSSLSPTSLVII
FT                                HLSMDYYYSFLLTYITHCSITSITSSSLEKQIRLVISPHII
FT                                VLFRLKH (in isoform D).
FT                                /FTId=VSP_004050.
FT   VARSPLIC     99    668       Missing (in isoform D).
FT                                /FTId=VSP_004051.
FT   VARSPLIC    280    328       Missing (in isoform E).
FT                                /FTId=VSP_004052.
SQ   SEQUENCE   668 AA;  74487 MW;  4507369593D1DD3B CRC64;

  Query Match             51.9%;  Score 54.5;  DB 1;  Length 668;
  Best Local Similarity   44.4%;  Pred. No. 9;
  Matches   16;  Conservative    4;  Mismatches    5;  Indels   11;  Gaps    1;

Qy          1 AXAEAAEKAAKYAA-----------EAAEKAAKAXA 25
              |  |||:|||: ||           |||:|||:  |
Db        497 AAEEAAQKAAEEAALKAAEEARLAEEAAQKAAEEAA 532


RESULT 6
CEA_CITFR
ID   CEA_CITFR      STANDARD;      PRT;   592 AA.
AC   P04480;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Colicin A.
GN   CAA.
OS   Citrobacter freundii.
OG   Plasmid ColA-CA31.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Citrobacter.
OX   NCBI_TaxID=546;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84036205; PubMed=6313941;
RA   Morlon J., Lloubes R., Varenne S., Chartier M., Lazdunski C.;
RT   "Complete nucleotide sequence of the structural gene for colicin A, a
RT   gene translated at non-uniform rate.";
RL   J. Mol. Biol. 170:271-285(1983).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88174422; PubMed=2832701;
RA   Morlon J., Chartier M., Bidaud M., Lazdunski C.;
RT   "The complete nucleotide sequence of the colicinogenic plasmid ColA.
RT   High extent of homology with ColE1.";
RL   Mol. Gen. Genet. 211:231-243(1988).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 389-592.
RX   MEDLINE=92235820; PubMed=1373773;
RA   Parker M.W., Postma J.P.M., Pattus F., Tucker A.D., Tsernoglou D.;
RT   "Refined structure of the pore-forming domain of colicin A at 2.4-A
RT   resolution.";
RL   J. Mol. Biol. 224:639-657(1992).
CC   -!- FUNCTION: This colicin is a channel-forming colicin. This class of
CC       transmembrane toxins depolarize the cytoplasmic membrane, leading
CC       to dissipation of cellular energy.
CC   -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC       AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC   -!- SIMILARITY: Belongs to the channel forming colicin family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X01008; CAA25503.1; -.
DR   EMBL; M37402; AAA72879.1; -.
DR   PIR; I40784; IKEBCA.
DR   PDB; 1COL; 15-JUL-93.
DR   InterPro; IPR000293; Channel_colicin.
DR   Pfam; PF01024; Colicin; 1.
DR   PRINTS; PR00280; CHANLCOLICIN.
DR   ProDom; PD002657; Channel_colicin; 1.
DR   PROSITE; PS00276; CHANNEL_COLICIN; 1.
KW   Antibiotic; Bacteriocin; Plasmid; Transmembrane; 3D-structure.
FT   TRANSMEM    528    548       POTENTIAL.
FT   TRANSMEM    555    575       POTENTIAL.
FT   HELIX       396    418
FT   HELIX       420    434
FT   TURN        435    435
FT   TURN        437    438
FT   HELIX       444    455
FT   TURN        456    456
FT   HELIX       458    460
FT   HELIX       464    475
FT   TURN        476    476
FT   HELIX       479    489
FT   HELIX       491    493
FT   TURN        496    497
FT   HELIX       498    515
FT   HELIX       519    530
FT   TURN        531    532
FT   HELIX       535    552
FT   HELIX       557    575
FT   TURN        577    577
FT   HELIX       578    586
FT   TURN        587    588
SQ   SEQUENCE   592 AA;  62992 MW;  B80FA1F52A8CFC5D CRC64;

  Query Match             51.4%;  Score 54;  DB 1;  Length 592;
  Best Local Similarity   56.5%;  Pred. No. 9.4;
  Matches   13;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKA 23
              | ||||||| :  || ||:  :|
Db        364 AAAEAAEKARQRQAEEAERQRQA 386


RESULT 7
HS88_NEUCR
```

```
ID   HS88_NEUCR     STANDARD;      PRT;   707 AA.
AC   O74225;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Heat shock protein Hsp88.
GN   HSP88.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 431-436; 439-445; 536-549; 575-586;
RP   596-602 AND 689-697, AND CHARACTERIZATION.
RX   MEDLINE=98225221; PubMed=9556627;
RA   Plesofsky-Vig N., Brambl R.;
RT   "Characterization of an 88-kDa heat shock protein of Neurospora crassa
RT   that interacts with Hsp30.";
RL   J. Biol. Chem. 273:11335-11341(1998).
CC   -!- SUBUNIT: BINDS HSP30 INDEPENDENT OF TEMPERATURE OR SUBSTRATE.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- INDUCTION: By heat shock.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- SIMILARITY: Belongs to the heat shock protein 70 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF069523; AAC23862.1; -.
DR   InterPro; IPR001023; Hsp70.
DR   Pfam; PF00012; HSP70; 1.
DR   PRINTS; PR00301; HEATSHOCK70.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; FALSE_NEG.
DR   PROSITE; PS00329; HSP70_2; FALSE_NEG.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   Heat shock; ATP-binding.
SQ   SEQUENCE   707 AA;  78673 MW;  8B077E8CC08BB4C1 CRC64;

  Query Match             51.4%;  Score 54;  DB 1;  Length 707;
  Best Local Similarity   56.5%;  Pred. No. 11;
  Matches   13;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              |: ||: |: | |||||||:  |
Db        667 AKKAEEEARKAKEAAEKAAQEGA 689


RESULT 8
ANP_LIMFE
ID   ANP_LIMFE      STANDARD;      PRT;    97 AA.
AC   P09031;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   Antifreeze protein precursor (AFP).
OS   Limanda ferruginea (Yellowtail flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Pleuronectidae; Limanda.
OX   NCBI_TaxID=8258;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88029483; PubMed=3665937;
RA   Scott G.K., Davies P.L., Shears M.A., Fletcher G.L.;
RT   "Structural variations in the alanine-rich antifreeze proteins of the
RT   pleuronectinae.";
RL   Eur. J. Biochem. 168:629-633(1987).
CC   -!- FUNCTION: Antifreeze proteins lower the blood freezing point.
CC   -!- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
CC       ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06356; CAA29655.1; -.
DR   PIR; S02376; S02376.
DR   InterPro; IPR000104; Antifreeze_1.
DR   PRINTS; PR00308; ANTIFREEZEI.
KW   Antifreeze protein; Repeat; Signal.
FT   SIGNAL        1     23
FT   PROPEP       24     48       REMOVED BY A DIPEPTIDYLPEPTIDASE
FT                                 (PROBABLE).
FT   CHAIN        49     97       ANTIFREEZE PROTEIN.
SQ   SEQUENCE   97 AA;  8865 MW;  62AD582DF8E459B6 CRC64;

  Query Match             50.5%;  Score 53;  DB 1;  Length 97;
  Best Local Similarity   56.0%;  Pred. No. 2.9;
  Matches   14;  Conservative    1;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |  |  ||| ||: |  |||| |
Db         57 AAAATAAAAAKAAADTAAAAAKAAA 81


RESULT 9
RL19_RHIME
ID   RL19_RHIME     STANDARD;      PRT;   177 AA.
AC   Q92L39;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   50S ribosomal protein L19.
GN   RPLS OR R03246 OR SMC03863.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC   -!- FUNCTION: This protein is located at the 30S-50S ribosomal subunit
CC       interface and may play a role in the structure and function of the
CC       aminoacyl-tRNA binding site (By similarity).
CC   -!- SIMILARITY: Belongs to the L19P family of ribosomal proteins.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL591793; CAC47825.1; -.
DR   HAMAP; MF_00402; -; 1.
```

```
DR   InterPro; IPR001857; Ribosomal_L19.
DR   Pfam; PF01245; Ribosomal_L19; 1.
DR   PRINTS; PR00061; RIBOSOMALL19.
DR   ProDom; PD002979; Ribosomal_L19; 1.
DR   TIGRFAMs; TIGR01024; rplS_bact; 1.
DR   PROSITE; PS01015; RIBOSOMAL_L19; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   177 AA;  19255 MW;  1BD19D6561AB8F22 CRC64;

  Query Match             50.5%;  Score 53;  DB 1;  Length 177;
  Best Local Similarity   65.5%;  Pred. No. 4.7;
  Matches   19;  Conservative    1;  Mismatches    5;  Indels    4;  Gaps    2;

Qy          1 AXAEAAEKAAKYAAE--AAE--KAAKAXA 25
              | | ||||||  |||  |||  |||:| |
Db        148 AQALAAEKAAAEAAEAKAAEEAKAAEAAA 176


RESULT 10
LED8_DAUCA
ID   LED8_DAUCA     STANDARD;      PRT;   555 AA.
AC   P20075;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Embryonic protein DC-8.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC   Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Queen Anne's Lace;
RX   MEDLINE=89384429; PubMed=2571069;
RA   Franz G., Hatzopoulos P., Jones T.J., Krauss M., Sung Z.R.;
RT   "Molecular and genetic analysis of an embryonic gene, DC 8, from
RT   Daucus carota L.";
RL   Mol. Gen. Genet. 218:143-151(1989).
CC   -!- FUNCTION: May play a role in late embryogeny.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic, protein bodies, and cell walls
CC       of zygotic embryo and endosperm tissue.
CC   -!- SIMILARITY: Belongs to the LEA type 1 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X16131; CAA34258.2; -.
DR   PIR; S04909; S04909.
DR   InterPro; IPR004238; LEA.
DR   Pfam; PF02987; LEA; 6.
KW   Repeat.
FT   DOMAIN       97    391       17 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT       97    114       1.
FT   REPEAT      115    125       2.
FT   REPEAT      126    140       3.
FT   REPEAT      141    154       4.
FT   REPEAT      155    176       5.
FT   REPEAT      177    191       6.
FT   REPEAT      192    205       7.
FT   REPEAT      206    216       8.
FT   REPEAT      217    237       9.
FT   REPEAT      238    259       10.
FT   REPEAT      260    281       11.
FT   REPEAT      282    303       12.
FT   REPEAT      304    325       13.
FT   REPEAT      326    343       14.
FT   REPEAT      344    358       15.
FT   REPEAT      359    376       16.
FT   REPEAT      377    391       17.
SQ   SEQUENCE   555 AA;  60260 MW;  D15E8A30E51BD1AB CRC64;

  Query Match             50.5%;  Score 53;  DB 1;  Length 555;
  Best Local Similarity   57.1%;  Pred. No. 12;
  Matches   12;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AEAAEKAAKYAAEAAEKAAKA 23
              ||| ||  :|   ||:|||:|
Db        196 AEAKEKTGEYKDYAAQKAAEA 216


RESULT 11
H1G_STRPU
ID   H1G_STRPU      STANDARD;      PRT;   217 AA.
AC   P07796;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1-gamma, late.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87172742; PubMed=3031476;
RA   Knowles J.A., Lai Z.-C., Childs G.J.;
RT   "Isolation, characterization, and expression of the gene encoding the
RT   late histone subtype H1-gamma of the sea urchin Strongylocentrotus
RT   purpuratus.";
RL   Mol. Cell. Biol. 7:478-485(1987).
CC   -!- FUNCTION: Histones H1 are necessary for the condensation of
CC       nucleosome chains into higher order structures.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the histone H1/H5 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M16033; AAA30059.1; -.
DR   PIR; A26721; A26721.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   217 AA;  22658 MW;  C7251EED3413B185 CRC64;

  Query Match             49.5%;  Score 52;  DB 1;  Length 217;
  Best Local Similarity   56.5%;  Pred. No. 7.3;
  Matches   13;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKA 23
              | |: |:|||| | :||: | ||
Db        189 AAAKPAKKAAKPAKKAAKPAKKA 211


RESULT 12
TOLA_PSEAE
ID   TOLA_PSEAE     STANDARD;      PRT;   347 AA.
```

```
AC   P50600;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TolA protein.
GN   TOLA OR PA0971.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO;
RX   MEDLINE=97113525; PubMed=8955385;
RA   Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT   "Identification and characterization of the tolQRA genes of
RT   Pseudomonas aeruginosa.";
RL   J. Bacteriol. 178:7059-7068(1996).
RN   [2]
RP   REVISIONS TO N-TERMINUS.
RA   Duan K., Sokol P.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: Involved in the tonB-independent uptake of proteins (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC       (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39558; AAC44660.2; -.
DR   EMBL; AE004530; AAG04360.1; -.
DR   PIR; E83525; E83525.
DR   InterPro; IPR006260; TonB_C.
DR   TIGRFAMs; TIGR01352; tonB_Cterm; 1.
KW   Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW   Complete proteome.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     37       POTENTIAL.
FT   DOMAIN       38    347       PERIPLASMIC (POTENTIAL).
FT   DOMAIN      209    216       POLY-ALA.
SQ   SEQUENCE   347 AA;  37935 MW;  EEDD4B04AA095945 CRC64;

  Query Match             49.5%;  Score 52;  DB 1;  Length 347;
  Best Local Similarity   56.0%;  Pred. No. 11;
  Matches   14;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              |  :||| | | ||| |:| | | |
Db        171 AKKKAAEDAKKKAAEEAKKKAAAEA 195


RESULT 13
ANP8_MYOSC
ID   ANP8_MYOSC     STANDARD;      PRT;    45 AA.
AC   P04368;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   Antifreeze peptide SS-8.
OS   Myoxocephalus scorpius (Shorthorn sculpin) (Daddy sculpin).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
OC   Cottoidei; Cottidae; Myoxocephalus.
OX   NCBI_TaxID=8097;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=85285003; PubMed=4029130;
RA   Hew C.-L., Joshi S., Wang N.-C., Kao M.H., Ananthanarayanan V.S.;
RT   "Structures of shorthorn sculpin antifreeze polypeptides.";
RL   Eur. J. Biochem. 151:167-172(1985).
CC   -!- FUNCTION: Antifreeze proteins lower the blood freezing point.
CC   -!- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
CC       ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
DR   PIR; A05163; A05163.
KW   Antifreeze protein; Repeat.
FT   MOD_RES       1      1       BLOCKED.
FT   REPEAT        9     21
FT   REPEAT       22     33
FT   REPEAT       34     45
SQ   SEQUENCE   45 AA;  4006 MW;  260C0BCC663B6878 CRC64;

  Query Match             49.0%;  Score 51.5;  DB 1;  Length 45;
  Best Local Similarity   66.7%;  Pred. No. 2.3;
  Matches   16;  Conservative    1;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          3 AEAAEKAAKYAAEAAEK-AAKAXA 25
              | ||  ||| ||:|| | |||| |
Db         14 AAAAALAAKTAADAAAKAAAKAAA 37


RESULT 14
AB31_CHLRE
ID   AB31_CHLRE     STANDARD;      PRT;   495 AA.
AC   Q8S339;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Inner membrane ALBINO3-like protein 1, chloroplast precursor.
GN   ALB3.1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND ASSOCIATION
RP   WITH THE LHCII COMPLEX AND PSAE.
RC   STRAIN=CC-621;
RX   MEDLINE=22204449; PubMed=12215522;
RA   Bellafiore S., Ferris P., Naver H., Goehre V., Rochaix J.-D.;
RT   "Loss of Albino3 leads to the specific depletion of the
RT   light-harvesting system.";
RL   Plant Cell 14:2303-2314(2002).
CC   -!- FUNCTION: Required for the insertion of some light-harvesting
CC       complexes (LHC) proteins into the chloroplast thylakoid membrane.
CC       Essential for the assembly and activity of LHC I and II. Its
CC       function is probably partly distinct from that of ALB3.2.
CC   -!- SUBUNIT: Associates with the LHCII complex and with the psaE
CC       subunit of the LHCI complex.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC       thylakoid membrane.
CC   -!- SIMILARITY: Belongs to the OXA1/oxaA family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF492768; AAM11662.1; -.
DR   InterPro; IPR001708; 60kDa_innermeb.
DR   Pfam; PF02096; 60KD_IMP; 1.
KW   Chloroplast; Membrane; Inner membrane; Transit peptide; Transmembrane.
FT   TRANSIT       1      ?       CHLOROPLAST (POTENTIAL).
FT   CHAIN         ?    495       INNER MEMBRANE ALBINO3-LIKE PROTEIN 1.
FT   DOMAIN        ?     75       LUMENAL (POTENTIAL).
FT   TRANSMEM     76     96       POTENTIAL.
FT   DOMAIN       97    206       STROMAL (POTENTIAL).
FT   TRANSMEM    207    227       POTENTIAL.
FT   DOMAIN      228    273       LUMENAL (POTENTIAL).
FT   TRANSMEM    274    294       POTENTIAL.
FT   DOMAIN      295    317       STROMAL (POTENTIAL).
FT   TRANSMEM    318    338       POTENTIAL.
FT   DOMAIN      339    441       LUMENAL (POTENTIAL).
FT   TRANSMEM    442    462       POTENTIAL.
FT   DOMAIN      463    495       STROMAL (POTENTIAL).
FT   DOMAIN      400    461       ALA-RICH.
SQ   SEQUENCE   495 AA;  51628 MW;  A9CDBF2C044AE37E CRC64;

  Query Match             49.0%;  Score 51.5;  DB 1;  Length 495;
  Best Local Similarity   60.0%;  Pred. No. 16;
  Matches   15;  Conservative    3;  Mismatches    6;  Indels    1;  Gaps    1;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | || ||:| : | ||||:|| | |
Db        405 ARAE-AEEAVERAREAAEEAAAAAA 428


RESULT 15
PT1_STRCO
ID   PT1_STRCO      STANDARD;      PRT;   556 AA.
AC   Q9KZP1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)
DE   (Phosphotransferase system, enzyme I).
GN   PTSI OR SCO1391 OR SC1A8A.11.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
CC       sugar phosphotransferase system (PTS), a major carbohydrate active
CC       -transport system. Enzyme I transfers the phosphoryl group from
CC       phosphoenolpyruvate (PEP) to the phosphoryl carrier protein (HPr).
CC       Enzyme I is common to all PTS.
CC   -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine =
CC       pyruvate + protein N(pi)-phospho-L-histidine.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the PEP-utilizing enzyme family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL939108; CAB88887.1; -.
DR   HSSP; P22983; 1DIK.
DR   InterPro; IPR008731; PEP-utilisers_N.
DR   InterPro; IPR008279; PEP_mobile.
DR   InterPro; IPR006318; PEP_P_trans.
DR   InterPro; IPR000121; PEP_utilizers.
DR   Pfam; PF05524; PEP-utilisers_N; 1.
DR   Pfam; PF00391; PEP-utilizers; 1.
DR   Pfam; PF02896; PEP-utilizers_C; 1.
DR   PRINTS; PR01736; PHPHTRNFRASE.
DR   ProDom; PD000940; PEP_utilizers; 2.
DR   TIGRFAMs; TIGR01417; PTS_I_fam; 1.
DR   PROSITE; PS00742; PEP_ENZYMES_2; 1.
DR   PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
KW   Phosphotransferase system; Transferase; Kinase; Sugar transport;
KW   Phosphorylation; Complete proteome.
FT   ACT_SITE    186    186       TELE-PHOSPHOHISTIDINE INTERMEDIATE
FT                                (BY SIMILARITY).
FT   MOD_RES     186    186       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   556 AA;  57271 MW;  7CC6F60B630C7CE7 CRC64;

  Query Match             49.0%;  Score 51.5;  DB 1;  Length 556;
  Best Local Similarity   68.2%;  Pred. No. 18;
  Matches   15;  Conservative    1;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          5 AAEKAAKYAAEAAE-KAAKAXA 25
              | || |: |||||| ||| | |
Db        228 AEEKKARLAAEAAERKAALAAA 249


Search completed: April 20, 2004, 21:59:34
Job time : 12 secs
```

OM protein - protein search, using sw model

Run on: April 20, 2004, 21:56:44 ; Search time 39 Seconds
(without alignments)
202.255 Million cell updates/sec

Title: US-10-019-482-1
Perfect score: 105
Sequence: 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	69	65.7	485	10	Q8RXD0	Q8rxd0 arabidopsis
2	69	65.7	924	10	Q9SU08	Q9su08 arabidopsis
3	64	61.0	755	10	Q9FP71	Q9fp71 oryza sativ
4	62	59.0	168	16	O69907	O69907 streptomyce
5	60	57.1	899	3	Q8NIZ0	Q8niz0 neurospora
6	59	56.2	413	16	Q83SA1	Q83sa1 shigella fl
7	59	56.2	421	16	Q8FJT1	Q8fjt1 escherichia
8	58	55.2	347	16	Q9RKL9	Q9rkl9 streptomyce
9	58	55.2	593	16	Q8ZNE5	Q8zne5 salmonella
10	57.5	54.8	711	4	Q7Z3F5	Q7z3f5 homo sapien
11	57.5	54.8	757	4	Q14234	Q14234 homo sapien
12	57	54.3	177	16	Q9AB65	Q9ab65 caulobacter
13	57	54.3	997	5	Q9W2J2	Q9w2j2 drosophila
14	57	54.3	1020	5	Q86PC3	Q86pc3 drosophila
15	57	54.3	1069	5	Q86BG1	Q86bg1 drosophila
16	56.5	53.8	508	5	Q9VGD2	Q9vgd2 drosophila
17	56.5	53.8	647	16	Q89IE4	Q89ie4 bradyrhizob
18	56.5	53.8	664	5	Q9VGD3	Q9vgd3 drosophila
19	56.5	53.8	694	5	Q8SWT7	Q8swt7 drosophila
20	56	53.3	92	13	Q9DF23	Q9df23 myoxocephal
21	56	53.3	124	16	Q7V6K8	Q7v6k8 prochloroco
22	56	53.3	387	5	O96113	O96113 plasmodium
23	56	53.3	496	2	Q8VQW6	Q8vqw6 azotobacter
24	56	53.3	508	3	Q875A8	Q875a8 podospora a
25	56	53.3	660	16	Q88YM9	Q88ym9 lactobacill
26	56	53.3	809	5	P90534	P90534 dictyosteli
27	56	53.3	2197	12	Q88876	Q88876 tomato ring
28	55.5	52.9	1171	3	Q9P3E2	Q9p3e2 neurospora
29	55	52.4	190	5	O15860	O15860 plasmodium
30	55	52.4	194	5	Q9N3U8	Q9n3u8 caenorhabdi
31	55	52.4	389	16	Q9CM70	Q9cm70 pasteurella
32	55	52.4	564	16	Q8YPL1	Q8ypl1 anabaena sp
33	55	52.4	909	10	Q9SU09	Q9su09 arabidopsis
34	54.5	51.9	638	16	Q89IE3	Q89ie3 bradyrhizob
35	54	51.4	101	2	Q9X342	Q9x342 bacillus an
36	54	51.4	119	4	Q8WU25	Q8wu25 homo sapien
37	54	51.4	298	2	Q8KYN2	Q8kyn2 bacillus an
38	54	51.4	302	5	Q25562	Q25562 naegleria g
39	54	51.4	432	16	Q7URR3	Q7urr3 rhodopirell
40	53.5	51.0	301	11	Q8BJK2	Q8bjk2 mus musculu
41	53.5	51.0	674	16	Q7WP24	Q7wp24 bordetella
42	53.5	51.0	674	16	Q7W1B9	Q7w1b9 bordetella
43	53.5	51.0	713	10	Q84NK5	Q84nk5 oryza sativ
44	53.5	51.0	1452	4	Q9H4A0	Q9h4a0 homo sapien
45	53.5	51.0	1512	4	Q9H4A1	Q9h4a1 homo sapien

ALIGNMENTS

RESULT 1
Q8RXD0
ID Q8RXD0 PRELIMINARY; PRT; 485 AA.
AC Q8RXD0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Auxilin-like protein (At4g12780).
GN AT4G12780.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY081334; AAL91223.1; -.
DR EMBL; BT009679; AAP81797.1; -.
DR InterPro; IPR001623; DnaJ_N.
DR SMART; SM00271; DnaJ; 1.
SQ SEQUENCE 485 AA; 54793 MW; 1054D1021DB52AD5 CRC64;

```
  Query Match             65.7%;  Score 69;  DB 10;  Length 485;
  Best Local Similarity   68.0%;  Pred. No. 1.1;
  Matches   17;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | | :|||| |||| ||| || |
Db        184 AAAGARDKAAKAAAEAREKAEKAAA 208


RESULT 2
Q9SU08
ID   Q9SU08      PRELIMINARY;      PRT;   924 AA.
AC   Q9SU08;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Auxilin-like protein.
GN   T20K18.130 OR AT4G12780.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
RA   Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL049640; CAB40995.1; -.
DR   EMBL; AL161534; CAB78320.1; -.
DR   PIR; T06636; T06636.
DR   InterPro; IPR001623; DnaJ_N.
DR   SMART; SM00271; DnaJ; 1.
SQ   SEQUENCE   924 AA;  102223 MW;  26E22C7C831EFF9B CRC64;

  Query Match             65.7%;  Score 69;  DB 10;  Length 924;
  Best Local Similarity   68.0%;  Pred. No. 2.1;
  Matches   17;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | | :|||| |||| ||| || |
Db        603 AAAGARDKAAKAAAEAREKAEKAAA 627


RESULT 3
Q9FP71
ID   Q9FP71      PRELIMINARY;      PRT;   755 AA.
AC   Q9FP71;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   P0458A05.18 protein (B1157F09.8 protein).
GN   P0458A05.18 OR B1157F09.8.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0458A05.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:B1157F09.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002870; BAB19409.1; -.
DR   EMBL; AP003207; BAB64083.1; -.
DR   Gramene; Q9FP71; -.
DR   InterPro; IPR001623; DnaJ_N.
SQ   SEQUENCE   755 AA;  83969 MW;  9B4D034850116493 CRC64;

  Query Match             61.0%;  Score 64;  DB 10;  Length 755;
  Best Local Similarity   54.5%;  Pred. No. 7;
  Matches   18;  Conservative    2;  Mismatches    5;  Indels    8;  Gaps    1;

Qy          1 AXAEAAEKAAKYAA--------EAAEKAAKAXA 25
              | ||| |:||| ||         || |:|||| |
Db        398 AAAEARERAAKAAAEAKERVAEEARERAAKAAA 430


RESULT 4
O69907
ID   O69907      PRELIMINARY;      PRT;   168 AA.
AC   O69907;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein SCO5619.
GN   SCO5619 OR SC2E1.36.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL939124; CAA19411.1; -.
DR   PIR; T34804; T34804.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   168 AA;  17934 MW;  72063B195040BD6E CRC64;

  Query Match             59.0%;  Score 62;  DB 16;  Length 168;
  Best Local Similarity   65.2%;  Pred. No. 2.7;
  Matches   15;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKA 23
              | |:||||| : || ||:||| |
Db        106 AEAKAAEKAERAAARAAKKAAPA 128


RESULT 5
Q8NIZ0
ID   Q8NIZ0      PRELIMINARY;      PRT;   899 AA.
AC   Q8NIZ0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Related to kinetoplast-associated protein KAP.
```

```
GN   5F3.190.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL807371; CAD37020.1; -.
SQ   SEQUENCE   899 AA;  99309 MW;  5A110FCA4C09D8F9 CRC64;

  Query Match             57.1%;  Score 60;  DB 3;  Length 899;
  Best Local Similarity   73.7%;  Pred. No. 26;
  Matches   14;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 EAAEKAAKYAAEAAEKAAK 22
              :|||:||| ||| |||| |
Db        511 KAAEEAAKKAAEEAEKARK 529


RESULT 6
Q83SA1
ID   Q83SA1      PRELIMINARY;      PRT;   413 AA.
AC   Q83SA1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Membrane spanning protein, required for outer membrane integrity.
GN   TOLA OR SF0558 OR S0571.
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
DR   EMBL; AE015086; AAN42202.1; -.
DR   EMBL; AE016979; AAP16075.1; -.
KW   Complete proteome.
SQ   SEQUENCE   413 AA;  42355 MW;  93E10F2C5DE60DE8 CRC64;

  Query Match             56.2%;  Score 59;  DB 16;  Length 413;
  Best Local Similarity   60.0%;  Pred. No. 16;
  Matches   15;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |:|||:||| ||  |:| |:| |
Db        143 ADAKAAEEAAKKAAADAKKKAEAEA 167


RESULT 7
Q8FJT1
ID   Q8FJT1      PRELIMINARY;      PRT;   421 AA.
AC   Q8FJT1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TolA protein.
GN   TOLA OR C0818.
OS   Escherichia coli O6.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR   EMBL; AE016757; AAN79291.1; -.
KW   Complete proteome.
SQ   SEQUENCE   421 AA;  43184 MW;  DB296626F056D385 CRC64;

  Query Match             56.2%;  Score 59;  DB 16;  Length 421;
  Best Local Similarity   60.0%;  Pred. No. 16;
  Matches   15;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |:|||:||| ||  |:| |:| |
Db        151 ADAKAAEEAAKKAAADAKKKAEAEA 175


RESULT 8
Q9RKL9
ID   Q9RKL9      PRELIMINARY;      PRT;   347 AA.
AC   Q9RKL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Probable peptidase.
GN   SCO4108 OR SCD17.12.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Brown S.P., Harris D.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL939118; CAB56389.1; -.
DR   GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR002886; Peptidase_M37.
DR   Pfam; PF01551; Peptidase_M37; 1.
KW   Complete proteome.
SQ   SEQUENCE   347 AA;  35432 MW;  456DFC61B6C2FF0D CRC64;

  Query Match             55.2%;  Score 58;  DB 16;  Length 347;
  Best Local Similarity   68.2%;  Pred. No. 17;
  Matches   15;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAK 22
              | ||||  | | ||||| || |
Db        163 AAAEAAAAAEKKAAEAAAKAEK 184


RESULT 9
Q8ZNE5
ID   Q8ZNE5      PRELIMINARY;      PRT;   593 AA.
AC   Q8ZNE5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative von Willebrand factor, vWF type A domain.
GN   YFBK OR STM2315.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008803; AAL21216.1; -.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00092; vwa; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   593 AA;  64640 MW;  595CA58158968357 CRC64;

  Query Match             55.2%;  Score 58;  DB 16;  Length 593;
  Best Local Similarity   65.2%;  Pred. No. 30;
  Matches   15;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AEAAEKAAKYAAEAAEKAAKAXA 25
              ||||::||| ||| ||  | | |
Db         58 AEAAQQAAKKAAEYAEAKALADA 80


RESULT 10
Q7Z3F5
ID   Q7Z3F5      PRELIMINARY;      PRT;   711 AA.
AC   Q7Z3F5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein DKFZp686F06102.
GN   DKFZP686F06102.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Human fetal kidney;
RA   Poustka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX537939; CAD97910.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   711 AA;  61765 MW;  95B624A99B4A989B CRC64;

  Query Match             54.8%;  Score 57.5;  DB 4;  Length 711;
  Best Local Similarity   60.7%;  Pred. No. 41;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db        446 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 473


RESULT 11
Q14234
ID   Q14234      PRELIMINARY;      PRT;   757 AA.
AC   Q14234;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Elastin.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
DR   EMBL; M17282; AAC98395.1; -.
DR   EMBL; M16983; AAC98395.1; JOINED.
DR   EMBL; M17265; AAC98395.1; JOINED.
DR   EMBL; M17266; AAC98395.1; JOINED.
DR   EMBL; M17267; AAC98395.1; JOINED.
DR   EMBL; M17268; AAC98395.1; JOINED.
DR   EMBL; M17270; AAC98395.1; JOINED.
DR   EMBL; M17271; AAC98395.1; JOINED.
DR   EMBL; M17272; AAC98395.1; JOINED.
```

```
DR   EMBL; M17273; AAC98395.1; JOINED.
DR   EMBL; M17274; AAC98395.1; JOINED.
DR   EMBL; M17275; AAC98395.1; JOINED.
DR   EMBL; M17276; AAC98395.1; JOINED.
DR   EMBL; M17277; AAC98395.1; JOINED.
DR   EMBL; M17278; AAC98395.1; JOINED.
DR   EMBL; M17279; AAC98395.1; JOINED.
DR   EMBL; M17280; AAC98395.1; JOINED.
DR   EMBL; M17281; AAC98395.1; JOINED.
DR   GO; GO:0005578; C:extracellular matrix; NAS.
DR   GO; GO:0030023; F:extracellular matrix constituent conferring. . .; NAS.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   InterPro; IPR003979; tropoelastin.
DR   PRINTS; PR01500; TROPOELASTIN.
DR   PROSITE; PS00453; FKBP_PPIASE_1; 1.
SQ   SEQUENCE   757 AA;  66136 MW;  23B7FE5B8AF85CA8 CRC64;

  Query Match             54.8%;  Score 57.5;  DB 4;  Length 757;
  Best Local Similarity   60.7%;  Pred. No. 44;
  Matches   17;  Conservative    1;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          1 AXAEAAEKAAKY-----AAEAAEKAAKA 23
              | | || |||||     || ||: ||||
Db        441 AQAAAAAKAAKYGVGTPAAAAAKAAAKA 468


RESULT 12
Q9AB65
ID   Q9AB65      PRELIMINARY;      PRT;   177 AA.
AC   Q9AB65;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   ATP synthase F0, B' subunit.
GN   CC0366.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005710; AAK22353.1; -.
DR   PIR; E87294; E87294.
DR   TIGR; CC0366; -.
DR   GO; GO:0015992; P:proton transport; IEA.
DR   InterPro; IPR002146; ATPsynt_B/B'sub.
DR   Pfam; PF00430; ATP-synt_B; 1.
KW   Complete proteome.
SQ   SEQUENCE   177 AA;  18465 MW;  6F0A2E32CC3D2912 CRC64;

  Query Match             54.3%;  Score 57;  DB 16;  Length 177;
  Best Local Similarity   60.0%;  Pred. No. 12;
  Matches   15;  Conservative    1;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | |||||: ||  |  ||| | | |
Db        110 ASAEAAERQAKEEAVLAEKLAAAEA 134


RESULT 13
Q9W2J2
ID   Q9W2J2      PRELIMINARY;      PRT;   997 AA.
AC   Q9W2J2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG18375 protein.
GN   CG18375.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
```

RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003453; AAF46699.2; -.
DR HSSP; Q13625; 1YCS.
DR FlyBase; FBgn0034606; CG18375.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 997 AA; 107821 MW; E712D400C2C4FD3D CRC64;

Query Match 54.3%; Score 57; DB 5; Length 997;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 15; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
| | || ||: ||||| :|| | |
Db 439 AAAAAAAAAAQAAAEAANQAATAAA 463

RESULT 14
Q86PC3
ID Q86PC3 PRELIMINARY; PRT; 1020 AA.
AC Q86PC3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RE13301p.
GN CG18375.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=y;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BT003215; AAO24970.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS50002; SH3; 1.
SQ SEQUENCE 1020 AA; 110434 MW; 42A3AE30EC71787B CRC64;

Query Match 54.3%; Score 57; DB 5; Length 1020;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 15; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
| | || ||: ||||| :|| | |
Db 462 AAAAAAAAAAQAAAEAANQAATAAA 486

RESULT 15
Q86BG1
ID Q86BG1 PRELIMINARY; PRT; 1069 AA.
AC Q86BG1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG18375-PB.
GN CG18375.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,

```
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA   Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA   Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003453; AAO41341.1; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00023; ank; 2.
DR   Pfam; PF00018; SH3; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00248; ANK; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50002; SH3; 1.
SQ   SEQUENCE   1069 AA;  115158 MW;  BF102B0C044F80DA CRC64;

  Query Match             54.3%;  Score 57;  DB 5;  Length 1069;
  Best Local Similarity   60.0%;  Pred. No. 72;
  Matches   15;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 AXAEAAEKAAKYAAEAAEKAAKAXA 25
              | | ||  ||: ||||| :|| | |
Db        511 AAAAAAAAAAQAAAEAANQAATAAA 535


Search completed: April 20, 2004, 22:00:27
Job time : 41 secs
```